

# **Quality Assessment of After Action Reports: Findings and Recommendations**

Sneha Patel, MSW, MPH

Commissioned by the National Academies of Sciences, Engineering, and Medicine,  
Committee on Evidence-Based Practices for Public Health Emergency Preparedness and Response

**Date: November 25, 2019**

## Contents

<b>1. Introduction</b>	3
<b>2. Method</b>	3
<b>3. Findings</b>	5
<b>3.1 After Action Report Characteristics</b>	5
<b>3.2 Quality Assessment Results (N=38)</b>	5
<b>4. Discussion</b>	6
<b>5. Limitations</b>	9
<b>6. Conclusion</b>	10
<b>7. References</b>	10
<b>Appendix A: ECDC 11-item tool for assessing AAR methodological rigor</b>	11
<b>Appendix B: Detailed Quality Assessment Results</b>	13

## 1. Introduction

Public health agencies typically conduct after action reviews following a public health emergency or simulated emergency (exercise) in an effort to identify lessons learned, strengths and weaknesses, and subsequently improve emergency preparedness and response capabilities.<sup>1</sup> In the United States, after action reports (AARs) are formally required by several agencies and organizations that fund or oversee aspects of public health and healthcare emergency preparedness and response, including the Assistant Secretary for Preparedness and Response (ASPR), and the Centers for Disease Control and Prevention (CDC).<sup>2</sup> While FEMA's Homeland Security Exercise and Evaluation Program (HSEEP) provides general guidance regarding the principles of capability-based evaluation and after action report templates, there is no standardized approach to conducting after action reviews or documenting findings.<sup>1, 3</sup> Furthermore, barriers to insightful AARs include concerns about liability, political response, constraints on staff time, and lack of experience and analytical skills.<sup>4</sup> Therefore, AARs vary with regard to methodological rigor, level of detail included, and validity.

This issue was encountered during a synthesis of AAR findings conducted to inform two papers commissioned by The National Academies of Sciences, Engineering, and Medicine, Committee on Evidence-Based Practices for Public Health Emergency Preparedness and Response entitled, *"Information Sharing with Technical Audiences: Findings from After Action Reports and Case Reports"* and *"Public Health Emergency Operations Coordination: Findings from After Action Reports and Case Reports."*<sup>5,6</sup> To better understand the methodological rigor of the AARs included in the analysis of these two papers, The Committee commissioned a quality assessment of the AARs. This report presents findings from the quality assessment, which was conducted through application of the 11-item appraisal tool designed by the European Centre for Disease Prevention and Control (ECDC). Recommendations based on the author's research expertise and practical experience as a former Director of Evaluation in a major metropolitan local public health department are also discussed.

## 2. Method

The Committee identified gray literature published by relevant domestic and international organizations and agencies related to information sharing with technical audiences and public health emergency operations coordination. This included Association of Public Health Laboratories (APHL), Assistant Secretary for Preparedness and Response (ASPR), the Association of State and Territorial Health Officials (ASTHO), Centers for Disease Control and Prevention (CDC), Center for Health Security, Council of State and Territorial Epidemiologists (CSTE), European Centre Disease Prevention and Control (ECDC), Disaster Information Management Research Center at the National Library of Medicine at the National Institutes of Health (NLM/NIH), Department of Homeland Security (DHS), Federal Emergency Management Agency (FEMA), US Government Accountability Office (GAO), National Association of County and City Health Officials (NACCHO), National Center for Disaster Medicine and Public Health (NCDMPH), Preparedness and Emergency Response Centers (PERRC), Public Health Canada, Public Health England, RAND Corporation, and the World Health Organization (WHO). Additionally, the committee obtained 370 after-action reports published from 2009 to 2019 from the Homeland Security Digital Library (HSDL).

In addition to the online search, the Committee proactively solicited reports, both published and unpublished, through a request for documents. The reports were solicited through internal listservs at the National Academies, as well as through external mechanisms. An online request was published on the committee's study webpage, and the Board on Health Sciences Policy distributed the call for reports through the Forum on Medical and Public Health Preparedness for Disasters and Emergencies and the Disaster Science Action Collaborative. Staff contacted CDC, the study sponsor, for document suggestions, and also had them disseminate the announcement to their networks, and particularly the former PERRCs and PERLCs networks. Additionally, staff sent targeted emails PHEPR practitioner associations (e.g., NACCHO and ASTHO) and disaster science organizations (e.g., DR2, NCDMPH, and ASPPH). Submissions were accepted through March 8, 2019. This proved to be an effective way to collect after action reports (AARs), theses, and white papers. Reports were further prioritized for review using a Sorting Tool developed with input from the Committee. Additional details regarding the AAR prioritization process can be found in the previously mentioned reports.<sup>5,6</sup> A total of 38 AARs were prioritized for analysis and are the subject of this quality assessment report.

The European Centre for Disease Prevention and Control (ECDC) 11-item appraisal tool for assessing AARs was applied to each of the 38 AARs to assess methodological rigor. The 11-item tool is designed to assist with the systematic documentation of methods used in AARs, compare validity, and potentially inform best practices for a standard template.<sup>7</sup> The tool includes the following criteria: *prolonged engagement with the subject of inquiry; use of theory; data selection; information sampling; multiple data sources; triangulation; negative case analysis; peer debriefing and support; respondent validation; clear report of methods of data collection and analysis (audit trail) and depth and insight* (see Appendix A for detailed tool guidance).

Matrices were created in an Excel spreadsheet to structure AAR characteristics (type of event, location, etc.) and calculate scores. Detailed findings and a list of AAR references are provided in Appendix B. Due to the general nature of some of the guidance related to the ECDC validation categories, efforts were made to standardize scoring of these categories for the purposes of this assessment:

- *Prolonged engagement*: Reports were given 2 points in the *prolonged engagement* category if they described large sample sizes, comprehensive data sources, and/or midcourse reviews. Reports received 1 point in this category if they mention multiple data sources but insufficient detail regarding participants.
- *Use of theory*: Reports that explicitly mentioned intentionally using a framework for evaluation (e.g. PHEP or HPP capabilities) were given the full 2 points for the *use of theory* category. Reports that did not explicitly mention use of a framework, but rather implied use based on how the report was organized received 1 point.
- *Data selection*: Only reports that explicitly provided rationales for *data selection* were given 2 points. Reports that provided rationales for some but not all subjects were given 1 point. Reports that implied purposive selection but did not mention any rationale received 0 points, thus potentially skewing scores to the lower end in this category.
- *Peer debriefing and support*: Reports received 2 points in this category if the AAR was written by an independent consultant. Reports reviewed by an external entity (e.g. state

agency) for further validation received 1 point as they do not fulfill the “other researchers or investigators” ECDC criterion, but demonstrate an attempt at peer validation.

### 3. Findings

#### 3.1 After Action Report Characteristics

The ECDC 11-item tool was applied to a total of 38 AARs. Approximately 61% of the reports were based on real events. Full scale and functional exercises accounted for 16% and 21% of the reports, respectively. One report was self-categorized as both a full scale and functional exercise AAR. Hazards and threats ranged from infectious diseases (H1N1, Ebola, Hepatitis A, etc.), natural disasters, and man-made disasters (oil spill, explosion, etc.). Incident years ranged from 2009 to 2017 in 20 states in the United States. AARs were all published either in the same year or the year following the real event or exercise. **Table 1** provides a summary of AAR characteristics.

**Table 1: After Action Report Characteristics**

Characteristics of AARs (N = 38)		
Type of Event	Real Event 61% (n=23)	
	Exercise 39% (n=15)	
	<ul style="list-style-type: none"> <li>• Full Scale 16% (n=6)</li> <li>• Functional 21% (n=8)</li> <li>• Full Scale/ Functional 3% (n=1)</li> </ul>	
Hazard/ Threats	Public health threat	Anthrax, Cyclospora, Ebola, H1N1, Hepatitis A, Influenza, Novel respiratory disease, TB
	Natural disasters	Earthquake, Flood, Hurricane, Nor'easter
	Other	Explosion, Loss of potable water, Oil spill, White powder
Incident Years	2009 - 2017	
Time to Publication	All AARs published either same or subsequent year after real event/ exercise	
Location	CA, CO, CT, DE, FL, IL, LA, MA, ME, MN, MT, NH, NJ, NY, OH, OK, OR, TX, WA, WI	

#### 3.2 Quality Assessment Results (N=38)

Overall, application of the ECDC 11-item tool yielded low scores. The average total score for the 38 reports was 5 out of a maximum of 22 points. **Table 2** provides a breakdown of total points scored. A majority of AARs (58%; n=22) scored 4 points or lower, followed by 26% (n=10) in the 5-9 point range. Thirteen percent (n=5) of AARs scored 10-14 points, and only one report scored in the 15-19 point range. The highest total score was 16 out of 22.

Notably, two of the top three highest scoring AARs were written by consultants. Additionally, AARs based

**Table 2: Breakdown of total ECDC points scored**

Total points scored (0 min – 22 max)	% and # of AARs (N=38)
20-22 points	0%; n=0
15-19 points	3%; n=1
10-14 points	13%; n=5
5-9 points	26%; n=10
0-4 points	58%; n=22

on real events scored higher on average (6.61 points out of 22) than AARs based on exercises (3 out of 22 points).

**Table 3** below details average, min, median, and max scores by each of the 11 ECDC validation categories. The only two categories with an average score greater than 1 were *use of theory* and *depth and insight*. Most AARs (71%; n=27) did not provide sufficient detail to determine whether *prolonged engagement* occurred. On average, reports received the highest score in the *use of theory* category (1.63 out of 2). The vast majority (92%) of AARs did not provide a rationale for *data selection*. More than half of the AARs did not mention *information sampling* (61%; n=23) or *multiple data sources* (63%; n=24), making difficult to ascertain the appropriateness of the sample or sources used to inform AAR findings. Commonly cited methods included hotwashes, after action debriefs, after action meetings/conferences, document review, exercise evaluation guides and exercise participant feedback forms. Some AARs also mentioned use of surveys, focus groups, and in-depth interviews. Only one AAR explicitly described triangulation of results, while 36% (n=14) implied triangulation as findings were discussed based on one evidence base.

None of the AARs mention *negative case analysis* or *respondent validation*. Only three AARs described *peer debriefing and support*; two of which were written by consultants and one validated regional findings at the state level. Only one AAR described a comprehensive and thorough *audit trail*. Although AARs scored low overall, *depth and insight* was found to be of higher average quality than the other categories. While there is potential for bias without knowing key details about the sample and methods, it is possible that AAR authors are prioritizing writing more detailed analyses without considering the importance of also detailing methods.

**Table 3: Scores by ECDC validation category**

Validity category (N=38)	Mean Score	Min	Median	Max
1. Prolonged engagement	0.47 / 2	0	0	2
2. Use of theory	1.63 / 2	0	2	2
3. Data selection	0.13 / 2	0	0	2
4. Information sampling	0.42 / 2	0	0	2
5. Multiple data sources	0.68 / 2	0	0	2
6. Triangulation	0.42 / 2	0	0	2
7. Negative case analysis	0.00 / 2	0	0	0
8. Peer debriefing and support	0.13 / 2	0	0	2
9. Respondent validation	0.00 / 2	0	0	0
10. Audit trail	0.26 / 2	0	0	2
11. Depth and insight	1.05 / 2	0	1	2
<b>TOTAL SCORE</b>	<b>5.18 / 22</b>			

## 4. Discussion

### Application of ECDC 11-item tool

The ECDC 11-item tool for assessing AAR methodological rigor proved useful for structuring the quality assessment of the AARs. The tool captures key criteria that enable a

standardized approach to assess the comprehensiveness of AAR processes, methods, and documentation practices. Some of the categories were more difficult to apply than others due to the general nature of the tool guidance. For instance, the *prolonged engagement*, *use of theory*, and *data selection* categories required additional definitions noted in the methods section above to ensure a standardized approach was used for scoring purposes. However, the flexibility of the scoring for these categories may allow for broader application of the tool as AARs improve in quality. It is also important to note that this flexibility may influence reliability of the tool if different users skew towards either more lenient or strict scoring when assessing specific validity categories. For the purposes of this quality assessment, the ECDC 11-item tool was both suitable and informative.

### Overall quality of AARs

Findings from the quality assessment indicate a significant need to improve both AAR processes and level of detail included in the reports themselves. It is unclear if AAR authors simply left out basic methodological information that was otherwise rigorous, or if the reports would have scored low even if the requisite categories had been included. Based on this author's experience, AAR methods sections may be succinct as target audiences (e.g. internal agency staff) are more interested in the analysis and concrete corrective actions resulting from the after action review. Therefore, methods sections may be briefer than they would otherwise be for a formal publication. Based on the overall low scores of the AARs that did include information about *information sampling* and *data sources*, however, it seems more likely that a need for significant improvement would have been identified even if more of the AARs had comprehensively addressed the ECDC validity categories.

### Exercise-based vs. real event AARs

As noted earlier, real event AARs scored higher than exercise-based AARs. It is possible that this is because agencies invest more time and effort into learning from real world experience than from exercises. While this may be reasonable as real events allow for lessons learned that are grounded in practical experience, public health agencies should also consider more rigorous evaluation processes for functional and full-scale exercises given the low frequency of real emergencies.

### Prolonged engagement and use of theory

*Prolonged and repeated engagement* of the processes and people involved in the after action process appears to be a gap based on the quality assessment findings. AARs that scored higher in this category had large sample sizes, multiple data sources, and thoughtful analyses. Limiting sources to one after action meeting with a large number of participants may impede validity of findings if findings are based on feedback from participants who spoke up in the review meeting. Therefore, further enhancing the nature of engagement may help improve confidence in AAR findings.

Additionally, although most AARs mention alignment of findings to capabilities with regard to *use of theory*, it is unclear if this is the most useful application of theory. For instance, the majority of AARs that explicitly mention alignment to capabilities draw verbatim from the standardized HSEEP AAR template. Therefore, the utility of this approach to public health agencies is unknown. It is possible that the HSEEP template language is used because AARs are required for funding. Further exploring how effective jurisdictions find the PHEP and HPP

capabilities approach specifically for organizing and tracking progress made on lessons learned is recommended.

#### *Data selection, information sampling, multiple data sources, and audit trail*

AARs were particularly limited in their description of *data selection* rationales. However, most reports implied the rationale. For instance, participants involved in exercises were likely also the hotwash participants and selected because they had firsthand experience with the capabilities that were being tested even if actual participant affiliations and roles were not explicitly stated in the report. AARs would benefit from additional detail related to the roles of the participants that were involved in providing feedback to allow for greater clarity around the reason for selection and representativeness of the sample. Similarly, with regard to *information sampling*, it is often unclear if the samples selected are representative of those who should have been included. Overall, most AARs did not provide sufficient detail to adequately assess the appropriateness of the sample. Use of *multiple data sources* also appears to be an area for improvement as most reports did not specify the data sources used and those that did generally failed to mention multiple examples of two methods as per the ECDC tool guidance. Only one AAR provided enough information for an *audit trail*. Again, it is unclear if this is because authors preferred to limit the length of the AAR, or if it was left out due to the lack of rigorous methods.

#### *Triangulation and negative case analysis*

Although reports that mentioned multiple data sources were scored as implying triangulation, it is unclear if this is accurate given the overall lack of rigor of the AARs. Therefore, it is worthwhile to further explore the methods employed by agencies to see if greater analytic capacity needs to be developed or if it is simply an issue of specifying that triangulation was employed as part of the analysis. Similarly, given the limited information provided about samples and methods, it is unclear if negative case analysis was considered and not documented, or if it was neglected altogether.

#### *Respondent validation, and peer debriefing and support*

*Respondent validation* was not mentioned at all in the AARs. However, based on this author's experience, it is possible that AARs were circulated internally for review, even if this was not explicitly documented in the AAR itself. Further inquiry may be warranted to determine if this is an issue of insufficient documentation or a gap in after action processes. *Peer debriefing and support* appears to be a concrete gap as AARs may be more likely to mention this step if it occurred to signify validated findings. To ensure an accurate review of findings, it is important to encourage jurisdictions to account for both of these validation steps in the after action review processes.

#### *Depth and insight*

Although AARs scored overwhelmingly low in the other validity categories, the *depth and insight* provided in the reports indicate that root cause analyses may have been more thorough despite the potentially limited data sources and methods used. There is, however, an unknown potential for bias given the limited information related to the other validity categories. Enhancing practices related to the 11 categories would build greater confidence in the *depth and insights* provided in AARs.



### Consultants

As previously mentioned, two of the top three scoring AARs were written by consultants. Although many agencies likely lack the resources to contract consultants, there may be a relationship between improved quality and use of external evaluators. Consultants may have the requisite expertise and capacity to conduct more rigorous evaluations and may help minimize potential for bias by providing an external lens. That said, realistically, it is unlikely that resource-constrained agencies will be able to outsource evaluations to consultants. Therefore, it is worthwhile to consider allocating greater resources for strengthening the evaluation expertise of agency staff to further improve the quality of internally written AARs.

### Equity considerations and inclusive processes

While considerations of equity are implicitly accounted for in the 11-item tool, it is important to call out the importance of representative samples to ensure an inclusive evaluation process. For example, the evaluation process should include staff of various levels as well as input from not only external partners, but also the communities that jurisdictions ultimately seek to serve. Excluding communities from the after action process is a missed opportunity to hear from diverse voices that may not necessarily be reflected by the demographic of ICS leadership or staff.

Additionally, limiting samples to ICS leadership may skew findings to a leadership perspective at the cost of including feedback from staff more directly engaged in response operations. Furthermore, it is important to explore ways to segment participants to encourage greater candidness in debriefings, hotwashes, or focus groups. Grouping leadership with general staff in feedback sessions may discourage staff from fully expressing their views if they have critiques related to how a response was handled by leadership. Some AARs mention asking ICS leadership to gather feedback from their sections and having the leadership serve as representatives in debriefs. This approach may not adequately address concerns within a specific section. Because many jurisdictions did not seem to conduct surveys, it is especially important to consider the influence of hierarchy in group feedback sessions. Furthermore, relying solely on survey data for this type of candid feedback may be a missed opportunity to obtain richer data.

To promote equitable evaluation processes, it is recommended that multiple methods of data collection are used with diverse samples and that negative case analysis is conducted to further boost validity of AAR findings.

## **5. Limitations**

Although the scope of the AARs assessed was focused on reports that were of specific interest to the Committee, findings are limited by the small sample size. Findings may also be limited by selection bias as only AARs publicly available or volunteered by jurisdictions were included in this review. It is possible that AARs considered too sensitive to post publicly could have differing methodological rigor. Additionally, due to the somewhat subjective nature of the scoring for specific ECDC validity categories, it is possible that another independent rater may score the same set of AARs either higher or lower. However, that issue is less important in the context of this quality assessment as the tool was applied in a standardized way across the AARs of interest to the Committee, allowing for comparison across reports.

## 6. Conclusion

Findings from this quality assessment indicate that AARs typically leave out the majority of important validity categories that could foster greater confidence in after action findings. Although the reasons behind the low scores are unclear, guidance aimed at improving after action methods and the level of detail included in after action report methods sections is recommended for both transparency and quality purposes. The field of emergency preparedness and response could also benefit from drawing from the multi-disciplinary nature of public health to apply more rigorous evaluation processes when assessing lessons learned from public health emergencies.

## 7. References

- <sup>1</sup>Davies, R., Vaughan, E., Fraser, G., Cook, R., Ciotti, M., & Suk, J. (2018). Enhancing Reporting of After-action Reviews of Public Health Emergencies to Strengthen Preparedness: A Literature Review and Methodology Appraisal. *Disaster Medicine and Public Health Preparedness*, 13, 618-625.
- <sup>2</sup>Office of the Assistant Secretary for Preparedness and Response (ASPR) and Centers for Disease Control and Prevention (CDC). (2017). *Hospital Preparedness Program (HPP) and Public Health Emergency Preparedness (PHEP) Cooperative Agreements 2017-2022*. Retrieved from [https://www.cdc.gov/cpr/readiness/00\\_docs/PHEP-Funding-CDC-RFA-TP17-1701.pdf](https://www.cdc.gov/cpr/readiness/00_docs/PHEP-Funding-CDC-RFA-TP17-1701.pdf)
- <sup>3</sup>FEMA. *HSEEP Policy and Guidance*. (2013). Retrieved from <https://preptoolkit.fema.gov/web/hseep-resources>
- <sup>4</sup>Stoto, M. A., Nelson, C., Piltch-Loeb, R., Mayigane, L. N., Copper, F., & Chungong, S. (2019). Getting the most from after action reviews to improve global health security. *Globalization and Health*, 15, 58.
- <sup>5</sup>Patel, S. (2019). *Information Sharing with Technical Audiences: Findings from After Action Reports and Case Reports* (Unpublished).
- <sup>6</sup>Patel, S. (2019). *Public Health Emergency Operations Coordination: Findings from After Action Reports and Case Reports* (Unpublished).
- <sup>7</sup>European Centre for Disease Prevention and Control (ECDC). (2018). *Best practice recommendations for conducting after-action reviews to enhance public health preparedness*. Stockholm, Sweden: ECDC.

## Appendix A: ECDC 11-item tool for assessing AAR methodological rigor\*

Validity category description	Reviewer guidance notes
<b>1. Prolonged engagement</b> with the subject of inquiry. Has the review included lengthy and perhaps repeated interviews with respondents, and/or days and weeks of engagement within a case report site or group?	<b>Fully met (2):</b> gives the sense that engagement with incident has been thorough and deep as a result of long or repeated interviews, large sample sizes, prolonged or multiple site visits or stages of engagement, etc. <b>Partially met (1):</b> AAR has engaged with the subject well but does not appear comprehensive. <b>Not met (0):</b> they had superficial engagement or it was unclear.
<b>2. Use of theory.</b> Has theory been used to guide sample selection, data collection and analysis, and to draw guide interpretive analysis?	Does the AAR specify theoretical models, frameworks or approached used to inform their work? Does this include any of those recorded in the theoretical grounding work.
<b>3. Data selection.</b> Has purposive selection been used to allow prior theory and initial assumptions to be tested or to examine 'average' or unusual experience?	<b>Fully met:</b> clear sample rationales are given, providing a clear sense they have deliberately and purposively interviewed/studied their subjects. <b>Partially met:</b> rationale for who they have interviewed is brief or superficial, lacks detail, making it unclear why they have chosen the sample they have, or why it is limited in the way it is - e.g. 'meetings with key entities at the national, regional and local level, including health trusts, county governors and municipalities.' This example tells us the organisations interviewed but not their roles. <b>Not met:</b> rationale unclear or not reported.
<b>4. Information sampling.</b> Has the review gathered views from a wide range of perspectives and respondents rather than letting one viewpoint (person, organisation or specialty) dominate? Does it sample from enough people, places, times, etc. to ensure the influence of these factors on the behaviour and views of those people providing information is minimised. Is sampling expanded in the light of early findings?	Who did they select? Irrespective of whether a sample rationale has been given above, does the AAR appear to have picked an appropriately diverse sample? <b>Fully met:</b> wide and varied sample perspectives gathered e.g. the 'who contributed to the report' list is large and diverse. <b>Partially met:</b> key detail of the sample is missing - for example, the number interviewed, participants' roles or affiliated organisations. <b>Not met:</b> who they have interviewed is unclear or not reported.
<b>5. Multiple data sources.</b> Does the review seek multiple information sources (documents, personal testimony, site visits) and collate multiple examples of each? For example, are duplicate formal interviews with all sampled staff undertaken? Does it use researcher observation and informal discussion; are interviews conducted with people of different roles and levels of seniority?	<b>Fully met:</b> three main methods (testimony, records/reports, and site visit) fully met unless site visit is not applicable - e.g. looking at the role of leadership in a response, would not necessarily need a site visit. Fully met can be two methods but multiple examples of two methods - e.g. focus group, plus in-depth interviews, plus document review. <b>Partially met:</b> two methods, commonly testimony and document review. <b>Not met:</b> one method only - e.g. document/data review without personal testimony.
<b>6. Triangulation.</b> Does the review look for patterns of convergence and divergence by comparing results across multiple sources of evidence (e.g. across interviewees, and between interview and other data), between researchers, and across different methodological approaches? Does it also include comparisons within data – e.g. comparing different interview accounts.	<b>Fully met:</b> used words similar to triangulation or described methods of formally comparing and contrasting insight between, and/or within, data sources - e.g. do CCTV accounts verify eye witness accounts. <b>Partially met:</b> collected multiple sources of data and do not state how they synthesised them but it is implicit that they did as they talk about one evidence base. <b>Not met:</b> not reported.

<b>7. Negative case analysis.</b> Does the review look for evidence that contradicts its initial findings, explanations and theory, and refine them in response to this evidence?	Looking for specific mention of 'negative case analysis' or 'deviant case analysis' or reference to very similar approach described left.
<b>8. Peer debriefing and support.</b> Does the AAR include a step where the findings and reports are reviewed by other researchers or investigators?	Looking for specific mention of this or reference to very similar approach. Includes public consultation.
<b>9. Respondent validation.</b> Review of findings and reports by respondents to check investigator interpretation of their input.	<b>Fully met</b> if respondents have validated/had the opportunity to comment on the report findings of their views. Must be respondents. Other commentators = peer debriefing and support.
<b>10.</b> Clear report of methods of data collection and analysis ( <b>audit trail</b> ). Has the review kept and reported a full record of activities available to others and presented a full account of how methods evolved and were applied?	<b>Fully met:</b> clear and comprehensive methodological detail giving sense their methods could be replicated independently. <b>Partially met:</b> methods are brief and somewhat superficial but they are at least documented. Similarly, if the report links to full methods elsewhere and hard to find. Fully met if the methods appear in an appendix or if there are links to another document that is easy to find.
<b>11. Depth and insight.</b> Has the AAR established the direct/indirect root causes and underlying contributory factors linked to errors, inaction or latent failures?	<b>Fully met:</b> the results clearly discuss root causes alongside and contributory factors throughout and in a systematic way. <b>Partially met:</b> some causal factors behind errors are discussed, but not throughout, or systematically. Patchy insight. <b>Not met:</b> recommendations/results seem superficial - e.g. largely describing what happened without insight into why or how.

\*European Centre for Disease Prevention and Control (ECDC). (2018). *Best practice recommendations for conducting after-action reviews to enhance public health preparedness*. Stockholm, Sweden: ECDC.

## **Appendix B: Detailed Quality Assessment Results**

Appendix B: Detailed Quality Assessment Results based on application of ECDC 11-item tool

			1. Prolonged engagement		2. Use of theory		3. Data selection		4. Information sampling		5. Multiple data sources		6. Triangulation		7. Negative case analysis		8. Peer debriefing & support		9. Respondent validation		10. Audit trail		11. Depth & insight		Total (Indiv report)		Real or Ex		Hazard		Loc.		Event/Ex Yr		Pub Yr		
Test Case	ID	AAR Reference	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		Q9		Q10		Q11														
EOC	630	Wood County Health District. (2017). <i>2017 Regional Functional/ Full-Scale Exercise After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Very brief, limited info; root causes not discussed	2	FE/FSE		PanFlu	OH	2017	2017					
EOC	631	Florida Department of Health. (2017). <i>2017 Statewide Hurricane Full Scale Exercise After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	Descriptive, but doesn't always get at root cause	3	FSE		Hurricane	FL	2017	2017					
EOC	632	City of Nashua. (2012) <i>October Nor'easter After Action Report</i> .	0	Unclear	2	mentions aligning objectives to capabilities		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	insufficient root cause analysis; superficial findings	2	Real		Nor'easter	NH	2012	2012					
EOC	633	Oklahoma Department of Emergency Management. (n.d.). <i>Earth Wind and Fire After Action Report/ Corrective Action Plan</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	insufficient analysis and detail provided to understand root causes	2	FSE		Natural threats	OK	2013	Missin g					
EOC	634	Logan County Health District. (2015). <i>2015 Logan County Health District Full Scale Exercise After Action Report/ Improvement Plan</i> .	0	Unclear	1	organized by capability		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	insufficient analysis and detail provided to understand root causes	1	FSE		POD	OH	2015	2015					
EOC	637	Contra Costa Health Services. (2012). <i>Chevron Richmond Refinery Fire of August 6, 2012 After Action Report Based on Medical/Health Debriefing Conducted September 10, 2012</i> .	0	Unclear, lists long list of debrief participants, but level of engagement unclear as only debrief and document review mentioned	0	Not mentioned		0	rationale not provided	1	missing roles, provides names and orgs	1	debrief and documents	1	implied	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	provides details, but difficult to follow how findings link to recs due to the way the report is organized (lack of theory)	4	Real		Refinery Fire	CA	2012	2012					
EOC	638	DuPage County Health Department (2009). <i>H1N1 After Action Report - Improvement Plan</i> .	0	Unclear	2	Assessed by capability and activity		0	Not mentioned	0	Unclear	1	hotwash, discussions, feedback forms - did not give fully met score because unclear levels of seniority or nature of "discussions"	1	implied	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	2	Thorough although difficult to follow due to report format	6	Real		H1N1	IL	2009	Missin g					
EOC	639	Governor's Office of Homeland Security & Emergency Preparedness. (2012). <i>Hurricane Isaac After Action Report &amp; Improvement Plan</i> .	0	Unclear	0	Not mentioned		0	Not mentioned	1	only lists agency names	0	after action co	0	Not mentioned	0	Not mentioned	1	Not other investigators, but regional findings were validated by state level	0	Not mentioned	0	Not mentioned	0	high level summary of multiple AACs - does not discuss root causes in detail, consistently, or systematically	2	Real		Hurricane	LA	2012	2012					
EOC	641	Multnomah County Health Department. (2009). <i>'Swine Flu Multico' Spring 2009 H1N1 Response After Action Report/ Improvement Plan</i> .	0	Unclear	1	organized by capability		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	2	Thoroughly ties observations to analysis and recommendations	3	Real		H1N1	OR	2009	2009					
EOC	642	Becker County Community Health. (2013). <i>People and Stuff HSEM Region 3 Logistics Exercise After Action Report/ Improvement Plan</i> .	0	Unclear	2	Mentions rationale for tying ex objs to core capabilities for eval purposes		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Brief analysis; unclear if tied to root causes	2	FE		Storm	MN	2013	2013					
Both	643	Chicago Department of Public Health, Illinois Department of Public Health, & Metropolitan Chicago Healthcare Council. (2011). <i>Illinois Hospitals Pediatric Full-Scale Exercise After Action Report</i> .	1	Mentions many evaluators and a hotwash, but unclear who the hotwash participants were	2	mentions aligning exercise objectives to capabilities		0	Not mentioned	0	Mentions hotwash, however, unclear if other methods were also used	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	Analysis provides varying level of detail, but considers findings from the multiple hospitals participating in the exercise. Could have provided more detail for some sections, but it's likely that the report was kept intentionally brief.	4	FSE		Earthquake Mass Casualty (Peds)	IL	2011	Missin g					
IS	644	Montana Department of Public Health and Human Services (2014). <i>Big Sky Push II Full Scale Exercise After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Analysis was very brief, and some corrective actions didn't seem to get to possible root cause	2	FSE		Influenza	MT	2014	2014					
Both	646	New Hampshire Department of Safety and Department of Health and Human Services. (2009). <i>Cities Ready Initiative Operation Rapid RX Full-Scale Exercise After Action Report</i> .	0	Unclear; only mentions hotwash and sometimes only 2 evaluators for 50 participants for multi-site exercise	2	mentions aligning exercise objectives to capabilities		0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	Patchy and inconsistent levels of insight	3	FSE		Anthrax	NH	2009	2009					

Appendix B: Detailed Quality Assessment Results based on application of ECDC 11-item tool

			1. Prolonged engagement		2. Use of theory		3. Data selection		4. Information sampling		5. Multiple data sources		6. Triangulation		7. Negative case analysis		8. Peer debriefing & support		9. Respondent validation		10. Audit trail		11. Depth & insight		Total (Indiv report)		Real or Ex		Hazard		Loc.		Event/Ex Yr		Pub Yr		
Test Case	ID	AAR Reference	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		Q9		Q10		Q11														
Both	647	New Hampshire Department of Safety and Department of Health and Human Services. (2009). <i>New Hampshire Spring 2009 H1N1 Response After Action Report/ Improvement Plan</i> .	1	mentions multiple data sources, but not level of engagement	2	Assessed by capability and activity	0	Not mentioned	0	Not mentioned	2	Mentions survey, focus group, phone calls, and doc review	1	unclear, but discussed as one evidence base	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Method unclear; only sources mentioned	2	Practical insights provided	8	Real		H1N1	NH	2009	2009						
Both	648	Minnesota Department of Health. (2013). <i>Operation Loon Call 2013 After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Missing "analysis" section for some objectives; does not provide sufficient detail to back up recs	2	FE		Explosion (terrorism)	MN	2013	2013						
Both	651	County of San Diego. (2018). <i>San Diego Hepatitis A Outbreak After Action Report</i> .	0	Unclear	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	primarily descriptive	0	Real		Hep A	CA	2017	2018						
Both	652	Boston Public Health Commission (2013). <i>2013 Boston Marathon ESF-8 Health &amp; Medical Planning, Response, &amp; Recovery Operations After-Action Report/Improvement Plan</i> .	2	appears comprehensive due to various data sources	2	mentions capabilities	0	Not mentioned	1	lists orgs, but not roles	2	emails, response logs, records, interviews, surveys	1	implied	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	insufficient detail	1	Analysis and recs could go deeper into root causes	10	Real		Explosion (terrorism)	MN	2013	2013						
Both	653	Massachusetts Emergency Management Agency, Massachusetts Department of Public Health, City of Boston, City of Cambridge, Town of Watertown, Massachusetts Bay Transportation Authority Transit Police Department, Massachusetts National Guard, & Massachusetts State Police (2014). <i>After Action Report for the Response to the 2013 Boston Marathon Bombings</i> .	2	appears comprehensive due to large sample size	2	aligns to capabilities and describes after action process	0	Not mentioned	1	conducted 100+ interviews with variety of stakeholders, but does not provide details of roles within agencies	1	interviews, document review	1	implied	0	Not mentioned	2	written by consultant	0	Not mentioned	1	mentions methods but not details	2	very detailed and thorough	12	Real		Explosion (terrorism)	MA	2013	2014						
Both	654	Buffalo Hospital & Wright County Public Health. (2013). <i>Buffalo Hospital Closed POD After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning objectives to capabilities	0	Not mentioned	0	unclear who participants were (provides only agency names and unclear who was involved in external evaluations)	1	hotwash, external evaluations	0	unclear	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	insufficient analysis and detail provided to understand root causes	3	FE		Anthrax	NY	2013	2013						
Both	656	Capitol Region Council of Governments (2017). <i>Ebola Virus Disease Full Scale Exercise After Action Report</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities	0	Not mentioned	0	Does not specify which partners were included	0	Not mentioned	0	unclear	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	unclear	1	inconsistent level of detail	3	FSE		Ebola	CT	2017	2017						
Both	657	Capitol Region Council of Governments (2016). <i>Ebola Virus Disease Functional Exercise After Action Report</i> .	0	Unclear	2	mentions aligning exercise objectives to capabilities	0	Not mentioned	1	missing roles, provides names and orgs	0	After action meeting	0	Unclear	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	unclear	2	detailed analysis	5	FE		Ebola	CT	2016	2016						
Both	658	Metropolitan Medical Response System Capitol Region Connecticut (2016). <i>CT Region 3 ESF-8 Ebola Preparedness &amp; Response After Action Report/ Improvement Plan</i> .	0	Unclear	2	thoughtfully describes how objectives were tied to capabilities in the AAR	0	Not mentioned	1	unclear who was in meetings, but survey was sent to multiple agencies	2	meetings, after action review meeting, survey	1	implied	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	insufficient detail to determine exactly what was done, but includes survey ques	2	provides detailed analysis and seems to address root causes	9	Real		Ebola	CT	2015	2016						
Both	660	Florida Department of Health (2010). <i>2010 Deepwater Horizon Oil Spill After Action Report/ Improvement Plan</i> .	0	methods may imply superficial engagement	2	describes use of an after action process	0	Not mentioned	0	leads were instructed to obtain written feedback from direct reports, potential for bias	1	written feedback and after action meeting	1	implied	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	vague	2	discusses root causes	7	Real		Oil Spill	FL	2010	2011						

Appendix B: Detailed Quality Assessment Results based on application of ECDC 11-item tool

Test Case	ID	AAR Reference	1. Prolonged engagement		2. Use of theory		3. Data selection		4. Information sampling		5. Multiple data sources		6. Triangulation		7. Negative case analysis		8. Peer debriefing & support		9. Respondent validation		10. Audit trail		11. Depth & insight		Total (Indiv report)				
			Comments		Comments		Comments		Comments		Comments		Comments		Comments		Comments		Comments		Comments		Comments		Real or Ex	Hazard	Loc	Event/Ex Yr	Pub Yr
Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11																			
Both	661	Multnomah County Health Department. (2010). <i>H1N1 Fall 2009 MultCo After Action Report/ Improvement Plan</i> .	2 comprehensive	2 ties to capabilities	2 provides rationales	1 does not detail who was included in hotwashes and debriefings, but survey was sent to various ICS sections	2 hotwashes, surveys, interviews, observation, debriefings	1 implied, but survey analyzed separately -unclear how findings were incorporated into narrative	0 Not mentioned	0 Not mentioned	0 Not mentioned	1 provides survey but few details about other methods	2 thorough	13	Real	H1N1	OR	2009	2010										
Both	662	Minnesota Department of Heath Department. (2014) <i>DOC FE Flash Floods 2014 After Action Report/ Improvement Plan</i> .	0 hotwash and feedback forms; unclear if level of engagement was sufficient due to lack of detail provided	2 PHEP capabilities used for EEGs	0 Does not provide detailed rationale	0 only states "staff" as participants, does not state roles or numbers	2 feedback form, hotwash, EEGs	1 implicit	0 Not mentioned	0 Not mentioned	0 Not mentioned	0 Not mentioned	1 Could provide more detail	6	FE	Flood	MN	2014	2014										
Both	663	New Hampshire Department of Health and Human Services and Department of Safety (2010). <i>New Hampshire H1N1 Response After Action Report/ Improvement Plan</i> .	2 comprehensive	2 mentions organizing by capability	0 Not mentioned	1 robust data sources, but unclear roles, number of participants, etc.	2 surveys, focus groups, interviews, after action meetings, and multiple other data sources	2 discusses findings from specific focus groups in relation to each other and surveys	0 Not mentioned	0 Not mentioned	0 Not mentioned	1 not comprehensive	2 Comprehensive and thorough	10	Real	H1N1	NH	2009-2010	2010										
Both	664	Delaware Division of Public Health (2010). <i>Novel H1N1 Influenza Delaware Response After Action Report/ Improvement Plan</i> .	2 mentions daily hotwashes and multiple data sources	2 mentions organizing by capability	0 Not mentioned	0 does not specify, but includes public as stakeholders that provided feedback	2 hotwashes, surveys, critiques, interviews	1 implied	0 Not mentioned	0 Not mentioned	0 Not mentioned	0 insufficient detail	2 thoughtful analysis	9	real	H1N1	DE	2009-2010	2010										
Both	665	Ohio Department of Health. (2010). <i>Fall 2009 H1N1 Response After Action Report – Improvement Plan</i> .	1 appears comprehensive due to variety of methods and data sources; however not clear who the participants were for all methods, therefore giving 1 point	2 mentions use of AAR-IP guidance (FEMA)	1 mentioned for some but not all participants	1 inconsistent level of detail	2 survey, debriefs, focus groups, 1-1 interface, phone, emails, presentations, documentation, etc	1 implied	0 Not mentioned	0 Not mentioned	0 Not mentioned	1 provides details about feedback during demob process, and survey	2 comprehensive	11	Real	H1N1	OH	2009-2010	2010										
Both	666	Tri-County Health Department. (2017). <i>Public Health Emergency Dispensing Exercise (PHEDEX) After Action Report and Improvement Plan</i> .	0 Unclear	2 Evaluates by capability	0 Not mentioned	1 missing roles, provides names and orgs	0 Not mentioned	0 Not mentioned	0 Not mentioned	0 Not mentioned	0 Not mentioned	2 root causes discussed in relation to hotwash findings, showing level of analysis that goes beyond just hotwash notes	5	FSE	Novel respiratory illness	CO	2017	Missing											
Both	667	Texas Department of State Health Services. (2018). <i>Hurricane Harvey Response After-Action Report</i> .	0 unclear	2 refers to prep domains and core capabilities	0 Does not provide detailed rationale	1 unclear who specifically participated, but info sessions are from across the state	0 information sessions	0 Not mentioned	0 Not mentioned	0 Not mentioned	0 Not mentioned	1 Inconsistent, patchy level of insight	4	Real	Hurricane	TX	2017	2018											
Both	668	Texas Department of State Health Services (2010). <i>Texas Department of State Health Services Response to the Novel H1N1 Pandemic Influenza After Action Report</i> .	2 describes comprehensive engagement	1 ties to capabilities	2 provides rationales	2 diverse sample	2 thoroughly described phased process with hot wash, focus groups, interviews, surveys, etc.	1 implied, but not explicitly stated	0 Not mentioned	2 written by consultant	0 Not mentioned	2 comprehensive	2 Comprehensive and thorough	16	Real	H1N1	TX	2009-2010	2010										



Appendix B: Detailed Quality Assessment Results based on application of ECDC 11-item tool

			1. Prolonged engagement		2. Use of theory		3. Data selection		4. Information sampling		5. Multiple data sources		6. Triangulation		7. Negative case analysis		8. Peer debriefing & support		9. Respondent validation		10. Audit trail		11. Depth & insight		Total (Indiv report)					Real or Ex		Hazard		Loc		Event/Ex Yr		Pub Yr	
Test Case	ID	AAR Reference	Q1		Q2		Q3		Q4		Q5		Q6		Q7		Q8		Q9		Q10		Q11																
Both	669	Wisconsin Hospital Emergency Preparedness Program. (2010). <i>After Action Report (AAR) for H1N1 Influenza</i> .	0	Unclear	1	mentions use of after action review	0	Not mentioned	1	mentions multiple after action conferences, but does not provide details	0	after action conferences	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	Patchy insight, but recs seem based on root causes	3	Real					H1N1	WI		2009-2010		2010			
Both	670	Wisconsin Division of Public Health. (2010). <i>2009 H1N1 Influenza Response After Action Report/ Improvement Plan</i> .	1	Describes intentional engagement of stakeholders, but unclear if sufficient	1	Briefly describes stratification by target capability and thematic analysis	0	Does not provide detailed rationale	1	Does not specify which partners were included	2	conferences, key informant interviews, debriefings, surveys	1	Implicit	0	Not mentioned	0	Not mentioned	0	Not mentioned	1	Brief methods	1	Systematically ties gaps to recs, but unclear if eval is primarily based on participant feedback, or accounts for a deeper root cause analysis	8	Real					H1N1	WI		2009		2010			
IS	672	Public Health – Seattle & King County. (2009) <i>H1N1 Influenza (Swine Flu) 2009 King County ESF-8 After Action Report</i> .	0	Unclear	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Brief analysis that is vague; corrective actions not tied to analyses	0	Real					H1N1	WA		2009		Missing			
IS	673	Blue Earth County Public Health. (2014). <i>Information Sharing for TB Contact Investigation After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning to capabilities	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	2	Clear analysis sections	4	Real					TB Contact Investigation	MN		2014		2014			
IS	674	Minnesota Department of Health. (2013). <i>White Powder Incident November 2013 After Action Report/ Improvement Plan</i> .	0	Unclear	2	mentions aligning to capabilities	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Very limited information provided	2	Real					White Powder	MN		2013		2014			
IS	675	Ramsey County Public Health. (2014). <i>Operation Communication Woes After Action Report/ Improvement Plan</i> .	0	Mentions 1 evaluator, a hotwash, and MSEI, unclear if sufficient level of engagement	2	mentions aligning to capabilities	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Very limited information provided	2	FE					Loss of potable water	MN		2014		2014			
IS	677	Maine Center for Disease Control and Prevention. (2010). <i>Maine CDC 2009 H1N1 Influenza Pandemic After Action Summary</i> .	2	although unclear sample size, mentions midcourse reviews, formal and informal meetings and stakeholder engagement	1	describes after action review	0	Does not provide specific rationale	1	Does not specify which partners were included	2	survey, meetings, after action debriefs	1	implied	0	Not mentioned	0	Not mentioned	0	Not mentioned	0	Not mentioned	2	Root causes discussed, sufficient level of detail provided to back up findings	9	Real					H1N1	ME		2009		2010			
AVG TOTAL FOR ALL REPORTS:			0.47		1.63		0.13		0.42		0.68		0.42		0.00		0.13		0.00		0.26		1.05		5.18														
MEDIAN:			0		2		0		0		0		0		0		0		0		0		1		4														
MIN:			0		0		0		0		0		0		0		0		0		0		0		0														
MAX:			2		2		2		2		2		2		0		2		0		2		2		16														