CHAPTER

PERSPECTIVE MATTERS: THE CHALLENGES OF PERFORMANCE MEASUREMENT IN WILDFIRE RESPONSE NETWORKS

5

Branda L. Nowell[1], Toddi A. Steelman[1,2], Anne-Lise K. Velez[1], Sherrie K. Godette[1]

[1]North Carolina State University, Raleigh, NC, United States; [2]University of Saskatchewan, Saskatoon, SK, Canada

CHAPTER OUTLINE

INTRODUCTION

Disaster incident response has long been recognized as complex and beyond the scope of resources of any single organization or agency to address single-handedly. This has spurred continued interest in how to accomplish a coordinated, coherent response among a myriad of organizations and agencies that become active during a response phase of a disaster (e.g., Comfort, 2007; Faas and colleagues, 2016; Kapucu, 2006a, 2006b; Nowell & Steelman, 2013, 2015; Steelman, Nowell, Bayoumi, & McCaffrey, 2014). We have found this collection of actors to be productively viewed as an *incident response network* (Nowell & Steelman, 2013; Steelman et al., 2014). A social network perspective provides additional insight for understanding the structure and functioning of this collection of actors.

Social Network Analysis of Disaster Response, Recovery, and Adaptation. http://dx.doi.org/10.1016/B978-0-12-805196-2.00005-4

When incident response is understood as a network response, it necessitates that we extend our focus beyond the performance of any single organization or agency during the incident and toward an understanding of the incident response stakeholders, their relationships, and the setting as a whole. Performance within networks can be conceptualized at four levels of analysis: performance of individuals, performance of organizations/agencies, performance of subgroups, and performance of the whole network. Research on response networks has shown that consideration of these levels is critical because performance at one level does not necessarily equate to performance at higher levels (Mandell & Keast, 2008). In fact, complex interdependencies within networks have been shown to result in high performance at one level while reducing performance at other levels (Mandell & Keast, 2008).

While performance measurement and evaluation is a well-established field, the science of evaluating performance within the constraints of social networks or organizations remains in its infancy (Kenis & Provan, 2009; Mandell & Keast, 2008; O'Toole, 2015). Further, many of the traditionally upheld pillars for what constitutes quality performance measures meet significant challenges when applied to evaluating response networks. Empirical research on the performance of networks in disaster contexts is especially sparse (Magsino, 2009). Understanding what we mean by performance in disaster response networks is essential to realizing the full value of what networks can accomplish.

In this chapter, we focus on performance of whole networks. Specifically, we offer insights and findings from efforts to operationalize performance of incident response networks associated with complex wildfire events occurring in the wildland–urban interface.[1] We first review the existing literature on network performance to identify key considerations and principles associated with evaluating performance in networks. Second, we describe how these principles were applied in a large multisite field study of wildfire incident response networks. Lastly, we empirically examine the relevance of pluralism in conceptualizing and measuring network performance. We conclude with a discussion of the implications of these findings for advancing the study of performance in networks.

ASSESSING PERFORMANCE IN NETWORKS: PRINCIPLES FROM THE LITERATURE

For our purposes, we view network performance as a multidimensional construct that concerns a network's achievement of process and outcome goals relative to the constraints of the context in which the network operates. We take this definition from the diverse public administration literature outside of disaster studies on how network performance can be measured (Kenis & Provan, 2009; Mandell & Keast, 2008; Turrini, Cristofoli, Frosini, & Nasi, 2010). In this manner, we build on Turrini and colleagues' (2010, p. 529) definition of network effectiveness:

> the effects, outcome, impacts, and benefits that are produced by the network as a whole and that can accrue to more than just the single member organizations in terms of increasing efficiency, client satisfaction, increased legitimacy, resource acquisition, and reduced costs.

[1]The wildland urban interface is the place where people and forests meet and often create vulnerabilities to hazardous situations in the face of a wildfire.

Understanding network function is essential to understanding network effectiveness and hence performance (Provan & Milward, 1995; Provan & Sebastian, 1998; Turrini et al., 2010). Functionality of the network is important because you need to know what you want the network to accomplish so that it can be measured against appropriate criteria for its given purpose. In thinking about the network from a functionalist outcome perspective, we need to know the *interdependent* actions that only the network can perform (Mandell & Keast, 2008). This is distinct from organizational functionality, which can be described as a function for which effective execution does not require coordination with other network actors outside of the command-and-control structure of one's own organizational structure (Nowell & Steelman, 2013). In effect, this means identifying those interdependent functions and then identifying how those functions can best be measured. It means being very clear about what constitutes performance, including the key tasks we want the network to accomplish as a whole.

Myrna Mandell and Robyn Keast (2008) suggest that in addition to thinking about the specific outcome functions of the given network, we need to consider the processes or relationships that enable these key interdependent functions to take place. This might include measures related to trust, the strength of relationships among members, reciprocity, and levels of participation among others. Finally, outcome and process functions can be affected by the context in which the network operates, so network context characteristics are also important features to consider (Kenis & Provan, 2009; Turrini et al., 2010). If there is agreement in looking at the outcome and process functionality of performance within an appropriately understood context, then the question arises about how to go about measuring these interdependent indicators of performance.

Keith Provan and Brinton Milward (1995) made the case more than 20 years ago that a multimeasure, multiperspective approach to measurement is essential for the most complete understanding of network effectiveness. Multiple viewpoints are important to account for in operationalizing network effectiveness because different constituencies will have different preferences in terms of what goals or outcomes matter to them (Kenis & Provan, 2009; Mandell & Keast, 2008). Given the multilevel nature of networks and the need for coordination and cooperation, it is important to understand perspectives at different levels of operations, which again means conceptualizing and measuring effectiveness in multiple ways and across multiple perspectives (Mandell & Keast, 2008).

Accordingly, a summary of recommendations from the existing literature on conceptualizing and operationalizing whole-network performance would include the following principles:

1. Network-level measures should reflect functional outcomes that require, or can be affected by, the interdependent actions of multiple actors and therefore cannot be achieved by a single entity working alone (Provan & Milward, 1995; Provan & Sebastian, 1998; Turrini et al., 2010).
2. Network performance should be conceptualized holistically rather than using single-indicator assessments. It should focus on the collaborative process as well as the collective outcomes (Koliba, 2014; Mandell & Keast, 2008).
3. Network performance measures should attend to both actor-level performance in contributing to the broader goals of the network as well as network-level performance in accomplishing its goals.
4. Network outcomes should be calibrated to be appropriate to the exogenous conditions within which the network is operating (Kenis & Provan, 2009; Raab, Mannak, & Cambré, 2015).
5. Assessments of network performance should consider the perceptions of all network members rather than single informants (Mandell & Keast, 2008).

MEASURING NETWORK PERFORMANCE DURING WILDFIRE DISASTERS

There is little research directed at understanding how to measure network performance in disasters. Calls have been made to explore whole disaster networks and their interactions (Comfort & Haase, 2006; Kapucu, 2005a,b, 2006a,b; Magsino, 2009), but there is little research on the topic, with the exception of Kapucu and Garayev (2014). While there is empirical interest in performance as it relates to the temporal connection between preparedness before a disaster and how that influences efficacy during a disaster (Kapucu, Arslan, & Collins, 2010; Nowell & Steelman, 2015), empirical measures for functional network performance during a disaster are lacking.

For this study, we sought to assess performance of incident response networks in the context of large-scale wildfire events in the wildland–urban interface. Wildfire events are growing in intensity and scale throughout the United States (Westerling, Hidalgo, Cayan, & Swetnam, 2006) and globally (Bowman et al., 2011; Flannigan et al., 2013; Moritz et al., 2014). Climate change, historical land management practices, and the significant expansion of human developments into previously uninhabited wildland areas have all contributed to an increasing severity and impact of wildfires on communities (Williams, 2013).

Large-scale wildfire events, particularly those that exist at the boundary of a wildland–urban interface, are generally coordinated through an incident management team working within the structure of the incident command system. The incident command system is an organization comprised of designated roles and principles that enable both human and physical resources from multiple agencies to become integrated into a unified organizational structure centrally coordinated by an incident commander and his/her incident management team (Irwin, 1989). Importantly, while these incident management teams have command over fire operations, they generally must work in lateral coordination with other local and state agencies for many other disaster response functions, such as evacuation, road closure, sheltering, and public information.

Managing a large-scale wildfire, like responding to most disasters, typically involves multiple jurisdictions with overlapping responsibilities and mandates (Fleming, McCartha, & Steelman, 2015; Moynihan, 2009; Steelman, Nowell, Bayoumi, & McCaffrey, 2012). When a large-scale disaster strikes, local responder agencies and organizations need to be integrated with extra-local federal disaster management teams under challenging conditions. A network perspective can be useful in conceptualizing this diverse array of actors as a collective entity or whole network (Provan, Fish, & Sydow, 2007). We label this collection of actors as the *incident response network*, which is defined as

> the collection of individuals, organizations, and agencies that have sustained involvement during the event who aim to serve the community in minimizing and coping with damages brought on by the disaster.
>
> **(Nowell & Steelman, 2013, p. 235)**

We focus specifically on the response conditions of the disaster, as opposed to the preparedness or recovery phases. The incident response network is activated to respond to the disaster and assist the community to minimize the disaster's impacts.

Our previous discussion implies that frameworks for measuring network performance in disasters need to be explicit about the specific populations, levels of analysis, phase of disaster, and the disaster context in which performance is to be measured (Nowell & Steelman, 2013). To measure network performance, we first have to ask ourselves what network we are measuring and for what purpose. Accordingly, in this study, we utilized the principles previously outlined to develop a set of whole-network performance indicators for use in evaluating performance during large-scale wildfire events. The performance measurement process is described next as it relates to each principle.

Principle 1: Network Level Measures Should Reflect Functional Outcomes That Cannot Be Achieved by a Single Organization, Agency, or Group Working Alone

At the onset of this project, it was not clear what types of outcomes were organizational or agency-related versus whole network. In other words, while many functional areas and objectives exist in incident response during wildfires, it was not clear which of these would require network or subgroup-level coordination. Therefore, the first stage in developing performance measures was gaining deeper insight into incident response from a network perspective. This involved field observations on seven wildfires in the American West as well as semistructured interviews with 24 key informants who had extensive experience in land management and wildfire response. These individuals represented diverse perspectives in incident management during wildfires and occupied multiple roles. Key informants were identified using a snowball sampling strategy to identify information-rich cases (Quinn, 2015) based on the investigators' more than 10 years of experience in the field. Participants had collective experience of 646 years on a total of 824 large-scale wildfires. These interviews included questions about what constituted an effective incident response.

Principle 2: Network Performance Should Be Conceptualized Holistically Rather Than By Using Single Indicator Assessments

To address this principle, transcripts from the key informant interviews were qualitatively analyzed to identify key themes in characterizing a high-performing incident response network on a wildfire. This analysis resulted in the development of 30 items, which were then reviewed by experts in the field for both clarity and comprehensiveness in representing incident response network performance. These 30 items were then distributed to incident responders on 21 complex wildfire incidents in Oregon, Washington, Idaho, and Montana in the summer of 2013. Exploratory factor analysis on the 30 items revealed five subscales: (1) coordination and fire response, comprising nine items (alpha = 0.90); (2) evacuation management, comprising six items (alpha = 0.91); (3) sheltering and mass care comprising five items (alpha = 0.91); (4) public information management, comprising four items (alpha = 0.84); and (5) road closure coordination, comprising three items (alpha = 0.88). Survey respondents were asked their level of agreement on a five-point scale for each item (see Table 5.1).

Principle 3: Network Performance Measures Should Attend to Both Actor-Level Performance in Contributing to the Broader Goals of the Network as Well as Network-Level Performance in Accomplishing Its Goals

In addition to the five network performance subscales, which capture overall performance of the incident response network, we also collected data evaluating the performance of individual actors as well as organizations and agencies within the network. These questions were guided and informed by our fieldwork and key informant interviews. For example, at the individual level, we utilized a social network roster questionnaire to ask how responders would rate the effectiveness of other responders in coordinating with the other individuals who are members of the incident response network. At the organization and agency level, we asked responders to rate the effectiveness of the incident management team as a whole in being responsive to *local responders*.[2] For this chapter, we focus exclusively on the five subscales that represented the whole-network performance measures.

[2]Local responders are those who reside in the place where the fire is occurring. They have formal responsibilities in law enforcement (sheriff, state police), emergency management, and fire management (volunteer fire departments and fire service).

Table 5.1 Network Performance Criteria Measures
Evacuation Network Performance: 6 Items (alpha = 0.91)
1. Cooperating agencies were able to use existing evacuation plans to quickly establish a coordinated evacuation strategy. **2.** Residents received timely notification of evacuation status using clear, preestablished language to distinguish between an evacuation warning and an evacuation notice. **3.** Evacuations were executed in a timely and orderly fashion. **4.** Cooperating agencies had a prepared plan for how reentry into evacuated areas would be coordinated. **5.** Trigger points for when evacuated areas would be opened for reentry were clearly communicated to the public. **6.** Reentry was carried out in an organized and orderly fashion.
Coordination and Response Network Performance: 9 Items (alpha = 0.90)
1. A coordinated set of fire management objectives were agreed upon among all affected jurisdictions. **2.** All concerned jurisdictions prioritized maintaining good communication across agencies. **3.** Credit for success and effort was shared among agencies during public meetings and media events. **4.** There was a general willingness across agencies to offer assistance to other agencies or jurisdictions. **5.** "Borrowed resources" were released in a timely fashion to minimize burden on the lending agency. **6.** Community values at risk from wildfire were readily identified. **7.** Efforts to protect community values were appropriate given available resources and risks to firefighter safety. **8.** The overall strategy taken in managing this fire was appropriate. **9.** Local resources were incorporated into the incident management operations.
Sheltering and Mass Care: 5 Items (alpha = 0.91)
1. Sheltering options were clearly communicated to evacuees. **2.** Adequate sheltering options were prepared to house evacuees. **3.** Donations for evacuees were well coordinated. **4.** Auxiliary care needs of evacuees (e.g., food, water, clothing, transportation, spiritual, or mental health assistance) were adequately provided for. **5.** Adequate sheltering options were made available to evacuate pets and livestock.
Public Information: 4 Items (alpha = 0.84)
1. Local resources were leveraged to ensure timely dissemination of public information. **2.** Public information was coordinated among cooperating agencies to ensure continuity of the message. **3.** Social media was used effectively to provide timely public updates concerning the status of the fire. **4.** A system for communication with the media was put in place to ensure timely dissemination of public information.
Road Closure: 3 Items (alpha = 0.88)
1. All cooperating and fire management agencies maintained a timely awareness of the status of road closures. **2.** Trigger points for making decisions about road closures were proactively communicated to the local community. **3.** A consistent message was provided to the public about the status of road closures.

Principle 4: Network Outcomes Should Be Calibrated to Be Appropriate to the Exogenous Conditions Within Which the Network is Operating

Patrick Kenis and Keith Provan (2009) argued for consideration of variation in exogenous factors when considering appropriate indicators of performance. These included the: (1) governance structure of the network, (2) whether the network was voluntary or mandated, and (3) the stage of development of the network. While important considerations in networks may vary on these dimensions, these exogenous factors had little bearing on our study as all networks were voluntary and were all at the same developmental stage. Similarly, the structure of the incident command system means all incident response networks possess the same basic governance structure. Network performance was measured to capture differences in complexity between incidents as well as to capture perceptions of all actors involved in the responder network. Our focus here was in examining differences between the 21 incidents we investigated as well as variation in evaluations by individuals and subgroups within the same incident.

Principle 5: Assessments of Network Performance Should Consider the Perceptions of All Network Members Rather Than Single Informants

In this study, we identified all agencies and organizations that had a sustained role in responding to each of our 21 wildfire incidents. Identification was done first through interviews with incident management team liaison officers for each of the 21 incidents. This initial list of names of engaged organizations was then validated and augmented by follow-up interviews with other central actors involved in the incident—generally the sheriff and/or county emergency manager. At each stage in identifying the network, we also asked informants to identify the individual who played a leadership role in coordinating their agencies' operations with the rest of the network. This process continued until no new network members were identified. These individuals represent the incident response network for each of the 21 incidents and, accordingly, became the sample for our performance evaluation questionnaire.

INVESTIGATING PLURALISM WITHIN INCIDENT RESPONSE NETWORK PERFORMANCE ASSESSMENTS

The prevailing conventional wisdom related to performance measurement suggests that we should look for differences among incidents. In other words, if we are to take networks seriously (e.g., O'Toole, 1997), we should be able to distinguish high- and low-performing incident response networks. Further, traditional performance measurement advises to prioritize precise, objective measures that are based on quantities that can be influenced or controlled by those being evaluated (Neely, Richards, Mills, Platts, & Bourne, 1997). Such a goal relies on a realist perspective of performance, assuming that there is an objective measure of performance to be uncovered that will be understood and rated by all actors relatively equivalently.

The assumptions underlying the realist perspective have been heavily criticized (e.g., Mandell & Keast, 2008), particularly in the context of complex problem domains in which the ultimate outcomes are impacted by a complex system of factors, many of which are outside of the control of the actors. Further, emphasis on easily quantifiable performance outcomes (i.e., reducing the number of acres burned) can lead to goal conflict among relevant actors. For example, an emphasis in national wildfire policy on measuring effectiveness of incident response based on the numbers of acres burned has been attributed to the escalation in wildfire risk we face today (Williams, 2013). This is because fire-adapted landscapes are ecologically designed to burn in order to renew themselves, and lack of fire on the land leads to an excess of woody fuels that increase fire intensity when the next fire occurs (Mutch, 1970). All of this indicates that, when working in complex problem domains, there are few clear-cut objective measures of effectiveness. Rather, any observed outcome would be judged on its effectiveness based on its consequences relative to alternative outcomes. This requires normative judgments to be made about what is valued. Given that different people may legitimately come to different conclusions in these value judgments due to different goal orientations, scholars have argued that understanding performance in networks requires accounting for pluralism within the network (Mandell & Keast, 2008; Turrini et al., 2010). Accordingly, if we heed the call to look at the diversity of individual perspectives, we need to look at the differences among individuals and stakeholder groups within a network and not simply at differences in performance between incidents.

The preceding discussion suggests that variation among actors within responder networks is just as important to understand as variation between networks in terms of evaluation of performance. However, consideration of pluralism in thinking about and measuring network performance has received scant empirical attention. In this study, we utilized a multilevel approach—individuals, subgroups, whole network—to investigate within- versus between-network variation on network performance measures. Within-network variation refers to differences in network performance evaluations across individuals or between subgroups within an incident response network. Between-network variation refers to variation in the averaged performance ratings for each of the 21 incidents that we studied. We then explored whether within-network variation in ratings of network performance appears to cluster at a subgroup level. Institutional affiliations (e.g., whether you represent the incident management team, or United States Forest Service, or county emergency response) tend to be strong identifiers in incident response to wildfire (Steelman et al., 2014). Accordingly, we operationalized subgroup differences in this study using institutional affiliations.

To build both the theory and empirical base for better understanding pluralism and its implications for assessing network performance, we focused on two research questions:

1. How much variation in performance assessments occurs within versus between networks?
2. To what extent is within-network variation associated with institutional affiliations?

METHODS

The data collected for this analysis were part of a larger study conducted by Nowell and Steelman (NSF # CMMI-1161755; JFS L12AC20571). The study took place during large wildfire events occurring in counties at risk for wildfire within the wildland–urban interface in Idaho, Montana, Oregon, and Washington, in summer 2013. The wildfire events considered in this study met two study requirements: (1) they were of sufficient complexity to require a national or regional incident management team; and (2) the wildfires occurred in the wildland–urban interface where they threatened human populations and infrastructure.

Twenty-one incidents within our four-state sample region met the criteria for inclusion in the survey resulting in performance data on 21 responder networks. We sent surveys containing the network performance measures described previously to all members of the incident response network for each of the 21 incidents. In total, we invited 885 individuals representing 369 organizations/agencies to participate in the survey. As summarized in Table 5.2, survey data were collected from 551 respondents (58% response rate) representing the United States Forest Service, incident management teams, county emergency responders, and other host land agencies (i.e., United States Bureau of Land Management, state forestry, etc.).

Table 5.2 Survey Response Rates by Subgroup

Group	Response Rate
Incident management teams	103 (93%)
County organizations	207 (45%)
Forest service districts	79 (62%)
Forest service supervisor's office	94 (69%)
Other host agencies	28 (51%)
Total	*511*
Total response rate	*58%*

Hierarchical linear modeling, also known as multilevel modeling, is a useful approach to testing relationships between variables at different levels of analysis. In this study, we employed hierarchical linear modeling because of its capacity to help disentangle the amount of variation that occurred in performance indicators at different levels of analysis, as well its ability to test hypotheses regarding the source of that variation (Hox, 2010).

RESULTS

BETWEEN-INCIDENTS NETWORK PERFORMANCE VARIATION

To answer our first research question, we ran a null model for each of the subscales of our network performance measures separately. A null model in hierarchical linear modeling allowed us to test whether there is significant variation both within and between response networks—while accounting for variation in the other levels. For each performance subscale, an interclass correlation was calculated to provide a breakdown of the percentage of variation that occurs at each level of analysis. Descriptive analysis on the overall variability on the network performance subscales (see Table 5.3) indicated a negative skew with scale averages ranging from 3.87 to 4.44 on a five-point scale. However, suitable variability was observed on all outcomes with standard deviations ranging around 1.0. Across responder networks, respondents tended to rate coordination and response network performance ($M = 4.44$) higher than all other network performance subscales.

When modeling the source of that variability, results overwhelmingly indicated that the greatest portion of variance was between respondents rather than between networks. In fact, only one performance dimension—evacuation—approached significant network level variation in the null models. This meant that the difference in the average rating of performance between our 21 wildfire incidents was not significant relative to the differences that existed among raters within incidents. In other words, the incidents were statistically indistinguishable from each other in terms of being higher or lower performing networks on these metrics once we accounted for within-network variation.

There were, however, some differences observed between network performance subscales (see Table 5.4). For evacuation performance ratings, 17% of the variance was between incidents while 83% was between individuals. For public information performance, only 5% of the variance was between

Table 5.3 Network Performance Subscale Descriptive Statistics

Network Performance Dimension	Mean	SD	Min	Max
Evacuation	3.99	0.91	1	5
Public information	4.34	0.77	1	5
Coordination and response	4.44	0.73	1	5
Road closure	4.13	0.95	1	5
Sheltering and mass care	4.00	0.91	1	5

Table 5.4 Hierarchical Linear Modeling Model—Performance Scores by Incident

Network Performance Dimension	Covariance Estimate of Intercept (Std. Error)	ICC	*p*-value
Evacuation	0.147 (0.076)^	0.174	.053
Public information	0.001(0.001)	0.048	.142
Coordination and response	0.001 (0.001)	0.034	.201
Road closure	0.026 (0.024)	0.028	.287
Sheltering and mass care	0.053(0.065)	0.063	.416

^p = <*.10.*

incidents while 95% was between individuals. The remaining outcome measures did not approach significance, meaning that the mean differences between incidents, after controlling for within-incident variation, was limited. This gives credence to the idea that perspectives within networks matter when conceptualizing performance within networks. The question then becomes, "what drives the within-network variation?"

NETWORK PERFORMANCE VARIATION WITHIN INCIDENTS

In this second phase of analysis, we investigated whether stakeholder affiliation (subgroups) helped to explain differences between individuals in their ratings of network performance (see Table 5.5). In other words, we sought to understand whether individuals representing the same stakeholder group evaluated the performance of their respective response network more similarly to one another. To address this question, stakeholder affiliation was added as a level-one fixed-effect predictor in the null models described in phase one. Here, we find that modeling stakeholder affiliation at level one generated a less than 3% change in variance explained. Interestingly, we found that for evacuation and sheltering, public information, road closures, and mass care network performance scores, incident management teams were the only group that had significant differences ($p < .05$) in ratings compared to other stakeholder groups. Overall, members of the incident management team tended to rate the effectiveness of the network more positively than did the other network actors. For mass care performance ratings, county response agencies were significantly more negative in their evaluations relative to other groups.

Table 5.5 Hierarchical Linear Modeling Model: Network Performance Scores by Respondent Sample Group

	Network Performance Subscales				
Institutional Subgroup	Evacuation	Public Information	Coordination and Response	Road Closures	Shelter and Mass Care
County responders	0.35 (0.30)	0.01 (0.04)	0.01 (0.04)	0.37 (0.22)	0.73 (0.37)*
USFS – District Office	0.29 (0.33)	0.05 (0.05)	0.03 (0.04)	0.42 (0.24)	0.43 (0.41)
USFS – National Forest Headquarters	0.16 (0.34)	0.04 (0.05)	0.02 (0.04)	0.17 (0.24)	–0.15 (0.42)
Incident management team	0.63 (0.32)*	0.110(0.05)**	0.07 (0.04)	0.64 (0.23)*	0.92 (0.42)*
Change in residual variance explained	*>3%*				

Referent class = non-USFS host agency.
Public information and coordination reported as absolute values due to reflected data transformation for normalization.
*p = *.05;* **p = *<.1.*

DISCUSSION

There is general consensus in the literature that disaster response is best understood as a network phenomenon and that the outcomes we care most about during disasters rely on information flow, coordinated action, and collective impact within that network (Comfort, 2007; Magsino, 2009; Mandell & Keast, 2008). Therefore, to assess performance during incident response requires that one be able to assess the performance of the network. Since performance criteria constitute value judgments about what matters (Simon, 1976), it is important to be clear about what we measure when we want to evaluate performance in networks. Challenges arise when measuring performance in networks because networks violate many of the assumptions related to how and what should be measured. For instance, conventionally we think measurements should be exact, objective, and based on characteristics under the control of the participants (Neely et al., 1997; Sandström & Carlsson, 2008). In contrast, networks take place in messy problem domains that are rarely fully under the control of the network (Provan & Kenis, 2008). Further, differences in perspective and position in the network matter, thereby creating difficulty in measuring a single, objective reality (Mandell & Keast, 2008).

Network scholars argue that efforts to understand performance in networks should both anticipate and attend to diverse perspectives among network members (e.g., Mandell & Keast, 2008; Turrini et al., 2010). Unfortunately, this advice stands in sharp contrast to conventional thinking about performance measurement, which prioritizes interrater reliability and would therefore dismiss variation across informants nested within the same incident as measurement error. Our findings are provocative on this front, offering empirical evidence that illuminates just how significant this issue may be when studying network performance in disasters.

As Mandell and Keast (2008) predicted, we found that responders within the same incident varied significantly in their perceptions of how well the network performed. In fact, less than 20% of the variation in ratings on any measure could be explained by differences between incidents. Interestingly, certain performance dimensions (e.g., public information and evacuation) appeared to distinguish between incidents better than others. To find strong differences between incidents using a multiinformant approach, two things must happen. First, members within networks must share a similar perspective in evaluating that performance dimension. Second, these shared assessments by some networks must be, on average, substantially more or less positive relative to other networks. In our study, these two conditions appeared to be better met in relation to the outcomes of public information and evacuation. However, the differences between incidents on these outcomes—after controlling for within-group variation—only approached a traditional level of significance ($p < .1$).

Perhaps this finding is not surprising if one stops to consider that incident response networks are situated in highly diverse settings. Disasters are complex events that necessitate a wide array of functions to be carried out by a myriad of actors with complementary resources and skills. Within these diverse settings, subgroup identities can become more pronounced (Faas et al., 2016). For example, stakeholder distinctions such as local government versus state or federal government can become more important (Fleming et al., 2015). Therefore, it is reasonable to suspect that individuals within the same subgroup may share more similar perspectives when evaluating an incident relative to others in the network. Surprisingly, however, this was largely not the case.

The perception of how key activities occurred in an incident appeared to be weakly tied to institutional affiliation (subgroup) or role during the incident. When we investigated this further, however, the only group that really made a difference was the incident management team. These team members, as a group, tended to rate the areas of public information, sheltering and mass care, evacuation, coordination/response, and road closures more positively compared to ratings provided by other groups. In the end, we did not find strong evidence of variation at the incident or the subgroup level.

PERFORMANCE MEASUREMENT IMPLICATIONS

Drawing from cognitive theory, we can think of three processes that can affect why and how different individuals might evaluate performance differently. The first suggests that *sensing*, which includes what kind of information one encounters, is important. Actors within an incident response network will have different experiences and access to different sets of information. Therefore, where one is positioned in the network may greatly affect how one perceives the network to have performed.

The second process potentially responsible for variation in perceived performance is *sense-making*, or how one interprets what one sees or experiences. The theory of sense-making asserts that two people exposed to exactly the same experience will likely attend to and interpret that experience differently based on differences in their individual past experience (Weick, Sutcliffe, & Obstfeld, 2005). This would contribute to variation in how someone views effectiveness relative to others.

The third process possibly implicated in different perceptions of performance is *assessing/evaluating*, which includes the kind of normative judgments one makes about the interpretation of the data. Judgments can be influenced by a number of factors including the degree of personal importance one assigns to a given outcome, differences in normative beliefs in what high performance means in a given setting, as well as personal or professional biases stemming from enlightened self-interest in the network being evaluated more or less positively.

This individual difference perspective may assist in understanding why incident management teams rated various elements of performance slightly higher than the other stakeholder groups. We can think of several explanations based on the insights in the previous paragraph. First, in terms of sensing, incident management teams are structurally located in a central position in the network, which may give them access to different information than other positions. This could lead to a different appraisal of the performance criteria. Second, in terms of sense-making, the incident management team will typically have more experience on large-scale wildfires relative to other local participants in the network. Consequently, they engage in a different type of sense-making than less-experienced participants. Third, in terms of assessing/evaluating, they have more experience, which allows them to evaluate performance on a wildfire. This may translate into the incident management team using a different scale upon which to calibrate their experience on a wildfire. For example, for a local responder, an event may be the worst incident that they have ever experienced. For an incident management team that deals with these events multiple times a year, this incident may not even rank in the top 20 of the worst events that they have seen. Alternatively, incident management teams may have an inherent—and potentially tacit—bias to view the incident as having gone more positively because incident management teams are the most centrally involved actors in coordinating activities within an incident response network. Therefore, they may be likely to perceive a stronger association between an evaluation of the network overall and an evaluation of them as an incident management team. Finally, more than any other subgroup, the incident management team engages in daily discussion about the incident as a group.

Research has shown that sense-making and decision-making in a group is unique and can trend toward more intense or extreme ideas (Moscovici & Zavalloni, 1969). In other words, an individual's assessment of an incident may become even more positive or more negative depending on the sentiment of the group as the incident is being discussed. As such, it could be that the relatively more positive sentiments of the incident management team members are rooted in group sense-making dynamics. These explanations and others warrant future research. Given the degree of pluralism that was evident in the 21 incidents studied in this chapter, it is clear that we need to better understand the factors that help to explain differences in evaluations between responders.

However, even if we were able to fully explain actor-level variation in evaluation ratings of network-level performance, we are still left with a quandary. If we are to consider disaster response as a network phenomenon, improving disaster response will require an ability to differentiate and learn from higher- and lower-performing networks. Pluralism within disaster response networks *is not measurement error*—it is an endemic reality to the phenomenon of networks in complex problem domains (Koliba, 2014; Mandell & Keast, 2008; Provan & Kenis, 2008). Our study suggests that significant pluralism exists among actors who comprise a disaster response network and this has implications for how we think about conceptualizing and measuring network performance in disasters. Therefore, we need tools for differentiating between networks in such a way that still does not discount or ignore this reality.

LIMITATIONS AND FUTURE RESEARCH

Findings from the current study are provocative, suggesting that evaluating the performance of whole networks may not be possible when pluralism exists. Data to support this premise are drawn from one specific type of network—incident response networks convened in reaction to large-scale wildfire events. As such, the present study is limited in its ability to draw conclusions concerning how much pluralism exists in other network contexts. Rather, we are content to assert only that pluralism *can* be significant in disaster response networks and therefore should be considered in any network performance measurement endeavor.

In addition, since our findings suggest that pluralism may not be explained away easily based on obvious stakeholder groupings and differences, this raises additional issues about what might explain within-incident variation. One explanation is that variation is clustered within alternative subgroups that we did not test. Alternative groupings such as local versus extra-local participants, professional affiliations (e.g., public information, law enforcement, fire management roles; see Nowell & Steelman, 2015), or individuals with greater or less experience with large wildfires may help us understand variation at the subgroup or individual level. An additional alternative explanation is that evaluation of network performance on many disasters is largely an individual-level phenomenon. Conceptually, we understand that within networks, different constituencies will have different experiences and will interpret these experiences through different filters and with different points of reference. Differences in how these criteria are experienced can lead to different evaluations of performance effectiveness. Unpacking these biases will be important to further conceptualizing how we better understand pluralism in evaluations of network performance.

We see two promising paths forward for future research. Operationalizing high performance in terms of the level of agreement among network members is one possibility. For example, Mandell and Keast (2008) have argued that high-performing networks develop shared identity, goals, and understandings across members that would, presumably, result in higher agreement in evaluating the network's performance. In this approach, it is the standard deviation rather than the average that is considered. The theory here is that, if network members do not agree, the network did not function well to create a shared point of reference and therefore would be evaluated as a lower-performing network. This offers a different approach compared to the functionalist performance approach that we embraced in this study.

Another potentially promising methodological path forward is to focus on extreme case comparison methodologies. Just because there was not substantial variation across incidents when all 21 events were considered does not mean that there were not a few incidents that distinguished themselves from the pack. In fact, it was clear from our data that most of the observed between-network variation was being driven by a few key problematic incidents in which there were markedly lower scores on performance across responders. A qualitative comparison of these extreme cases against cases characterized by corresponding high scores could lead to key insights about what factors distinguish higher- and lower-performing incidents.

With this research, we have just scratched the surface of what we need to know to evaluate network performance. In this chapter, we have identified a way forward in terms of approaching how we might view network performance measurement, some of the assumptions about plurality and its consequences for network measurement and demonstrated how this kind of analysis can be accomplished. Next steps include further efforts to overcome the challenges identified in our own work and looking at other disaster networks to see how and if pluralism matters in other domains. More research is needed to understand the determinants of within-incident variation and build practical guidelines for developing measures of network performance that avoid the pitfalls of the past and account for the plurality that is integral to complex problem domains. Without this progress, we will never realize the full value of what networks can accomplish or how to understand them.

ACKNOWLEDGMENTS

Data for this study were supported by funds from the National Science Foundation, the Joint Fire Science Program, and the USFS Northern Research Station. It was collected as part of a research initiative led by Dr. Branda Nowell (North Carolina State University) and Dr. Toddi Steelman (University of Saskatchewan). See firechasers.ncsu.edu for more information. Special thanks to members of the Fire Chasers Project team and all the fire and emergency personnel who contributed their time and insights in support of this project. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the US government. Mention of trade names or commercial products does not constitute their endorsement by the US government.