HSR Health Services Research

DOI: 10.1111/1475-6773.12216
METHODS ARTICLE

Assessing Organizational Change in Multisector Community Health Alliances

Jeffrey A. Alexander, Larry R. Hearld, and Yunfeng Shi

Objective. The purpose of this article was to identify some common organizational features of multisector health care alliances (MHCAs) and the analytic challenges presented by those characteristics in assessing organizational change.
Data Sources. Two rounds of an Internet-based survey of participants in 14 MHCAs.
Study Design. We highlight three analytic challenges that can arise when quantitatively studying the organizational characteristics of MHCAs—assessing change in MHCA organization, assessment of construct reliability, and aggregation of individual responses to reflect organizational characteristics. We illustrate these issues using a leadership effectiveness scale (12 items) validated in previous research and data from 14 MHCAs participating in the Robert Wood Johnson Foundation's Aligning Forces for Quality (AF4Q) program.
Findings. High levels of instability and turnover in MHCA membership create challenges in using survey data to study changes in key organizational characteristics of MHCAs. We offer several recommendations to diagnose the source and extent of these problems.
Key Words. Health care organizations and systems, organization theory, psychometrics

Multisector health care alliances (MHCAs) are voluntary organizations that bring together a diverse array of stakeholders (including physicians, hospitals, health insurers, employers, government agencies, and consumers) to work collaboratively on health-related issues in a community. These organizations, sometimes known as partnerships, collaboratives, or coalitions, may or may not be incorporated as distinct 501(c)3 entities and typically have a purpose that overlaps to varying degrees with those of the participating organizations. MHCAs offer potential for improving health outcomes at the community level through coordination of care and alignment of incentives among providers, payers, purchasers, and consumers of health care (Blumenthal 2012).

Increased attention to MHCAs as an organizational vehicle for health care reform has resulted in calls for more research to help policy makers and practitioners understand if and how these organizations can promote such improvements, particularly over the long term (Koza and Lewin 1998; Spekman et al. 1998).

In this context, there is increasing recognition that alliances themselves are not static but change and develop over time, and that simple cross-sectional evaluations do not adequately capture these dynamic properties (Florin et al. 2000; Zakocs and Edwards 2006). However, the nature of MHCAs presents a number of methodological challenges for researchers who wish to quantitatively study MHCA change and development. The purpose of this article was to identify some common organizational dynamics of MHCAs and the methods and metrics challenges presented by those characteristics. In so doing, we hope to make researchers more aware of challenges associated with studying collaborative forms of organizations such as MHCAs using quantitative methods, as well as introduce several potential analytic strategies to gauge their impact on MCHA change assessment. The discussion is also likely to be of interest to practitioners and program funders who are interested in promoting and improving the functioning of collaborative organizations through rigorous research and evaluation.

How Alliances Differ from More Traditional Organizations

The nature of MHCAs, relative to more traditionally structured organizations, gives rise to many of the methodological challenges associated with their evaluation. MHCAs operate on the basis of voluntary collaboration rather than hierarchical control. Their authority to set agendas, allocate resources, and resolve conflict is tenuous, deriving more from consent than from equity ownership or contractual authority (Huxham 1996; Alexander et al. 2001). This suggests that alliance members are more loosely bound to the organization and can leave without serious consequences to themselves should the alliance take an unacceptable position or threaten entrenched interests.

Alliances also consist of members with more variable levels of resource and effort commitment to the alliance and varying degrees of overlap between

---

Address correspondence to Jeffrey A. Alexander, Ph.D., Department of Health Management and Policy, The University of Michigan, 109 S. Observatory, Ann Arbor, MI 48109-2029; e-mail: jalexand@umich.edu. Larry R. Hearld, Ph.D., is with the Department of Health Services Administration, University of Alabama at Birmingham, Birmingham, AL. Yunfeng Shi, is with the Center for Health Care and Policy Research, The Pennsylvania State University, University Park, PA.

their own institutional goals and activities and those of an alliance (Okubo and Weidman 2000; Swain et al. 2001). Members walk a fine line between commitment to the alliance and its goals, on one hand, and those of their home organizations and primary job responsibilities, on the other (Gamm 1998; Huxham 1996; Sink, 1996; Zuckerman, Kaluzny, and Ricketts 1995). If these commitments are not aligned or change, this duality can provide considerable stress on the ability of alliances to accomplish their goals and sustain the alliance over the long term. These factors, individually and collectively, make alliances like MHCAs relatively fragile organizations, susceptible to considerable flux in membership and participation.

Three Methods Challenges in Assessing Alliance Organizational Characteristics

In the following sections, we describe the issue of changing alliance membership in more detail and highlight examples of three analytic challenges that can arise when quantitatively studying the organizational characteristics of MHCAs: (1) assessing change in MHCA organization; (2) assessment of construct reliability, and (3) aggregation of individual responses to reflect organizational characteristics. We illustrate these issues using a leadership effectiveness scale (12 items) validated in previous research (Alexander, Hearld, and Mittler 2011). Specifically, we compare changes in the sampling frames used between the two survey periods (changes in membership), compare the mean levels of leadership effectiveness across the two survey periods (assess change in MHCA), assess the measurement invariance of the leadership construct between the two survey periods (construct reliability), and test whether it is reasonable and appropriate to aggregate individual responses to the alliance (i.e., calculate intraclass coefficients and reliability statistics). Unlike discussions of response bias, which are commonly featured in the literature on assessing organizational change, our focus on potential issues stemming from changes in the population of individual participants in MHCAs has received less attention.

Sample and Data Collection

Our illustrations draw on data from 14 MHCAs participating in the Robert Wood Johnson Foundation's Aligning Forces for Quality (AF4Q) program.

The premise of AF4Q is that the greatest improvements in the quality of care can be achieved when aligning the efforts of key players in health care, including health care providers (physicians/physician groups, nurses, and hospitals), health care purchasers (employers and insurers), and health care consumers (patients), through multistakeholder alliances. Because these alliances were all participants in the AF4Q program, they operated under the same general vision of improving the quality of care in their respective communities and were expected to utilize the same general strategies to pursue that vision. These strategies focused on initiatives related to public reporting, quality improvement, consumer engagement, and equity associated with race and ethnicity. The operational aspects of these strategies (e.g., how the initiatives were implemented), however, were determined by the alliances and thus differed across the sites. The alliances also utilized different organizational structures (e.g., partnership among local organizations, single organization coordinating efforts of other organizations) to implement these strategies.

Data for assessing change in MHCA structures and processes were drawn from two rounds of an Internet-based survey of alliance participants. Participants were defined as persons who were members of alliance leadership or work groups, including task forces, steering committees and boards, and all staff members. The survey contained items related to participants' perceptions of alliance leadership, decision making, conflict resolution processes, participation cost and benefits, and alliance management. The survey was fielded in each location over a 4-week period. The first round of the survey was fielded from October 2008 to October 2009; the second round was fielded from October 2010 to December 2012. The first round yielded a response rate of 48.5 percent (range 30.5–76.5 percent), with 623 of a possible 1,283 respondents completing the survey. The second round yielded a response rate of 56.5 percent (range 41.5–78.9 percent), with 604 of a possible 1,069 respondents completing the survey.

Fluidity of Alliance Membership and Implications for Measurement and Analysis

Changing organizational membership is an issue for all organizations, but it presents a larger challenge for MHCAs by virtue of the fact that alliance members are synonymous with the alliance as an organizational entity and the relatively tenuous ties that these participants have to an alliance. Because many of the structural barriers to entry and exit do not exist in MHCAs, membership composition may change more often and sometimes more dramatically. Thus, the origin of many of the issues confronting researchers who study MHCA

change is the fluidity of MHCA membership over time. Traditional approaches to assessing organizational change emphasize differences in perceptions of organizational characteristics (i.e., leadership) by the same organizational members between two or more time periods. Alternatively, change at the organization level can be determined by taking the difference between aggregated responses of a common set of organizational members between two or more time points. Both approaches depend on having a panel of respondents available over time so that any observed change can be attributed to change in the organization or perceptions of its members, rather than changes in cohorts of individuals comprising the organization.

In the absence of stable, formal structures, MHCAs and their membership become virtually synonymous. Even if organizational participation in MHCAs remains stable, the individuals "at the table" may not. Significant shifts in MHCA participation raise important concerns about whether longitudinal assessments of alliances are tracking the same organizations over time or, alternatively, different organizations as their member composition changes. The scope and degree of the membership stability issue are illustrated in data from the 14 AF4Q alliances (Table 1). These data, based on the sampling frames provided by each alliance, suggest that an average of 43 percent of MHCA membership changed from the first administration of the survey to the second administration, 24 months later. The extent of this change ranged from a low of 20 percent to a high of 61 percent.

Three organizational processes contributed to the high level of membership fluctuation. First, members active in time 1 dropped out of the alliance by the time of the second round of the survey (61 percent of round 1 sampling frame, on average). Second, new members who were not active in the first round were added by the time of the second round survey (43 percent of round 2 sampling frame, on average). Third, the basic structure of the MHCA itself changed, resulting in elimination of certain committees, oversight groups, etc., and by extension, involvement of participants in those groups. Follow-up with several high turnover sites corroborated these dynamics. For example, one alliance leader noted "most change reflects the ebb and flow of people's jobs and lives (one person on the list died). Other differences from Round 1 to 2 reflect people being less engaged because either their work priorities or responsibilities shifted during the particular period. A few reflect a reconfiguration underway in how we approach consumer engagement."

The combination of these forces indicates significant "churning" in MHCA composition, resulting in an average of only 56 percent of the round 2 frame common to both rounds of the survey (range: 40–81 percent). As

Table 1: Alliance Survey Sampling Frames

Sites	*Round 1 Number in Sampling Frame*	*Round 2*						
		Number in Sampling Frame	*New Participants*	*% of Round 2*	*Panel Participants*	*% of Round 2*	*Dropouts*	*% of Round 1*
Cincinnati	77	106	60	56.60	46	43.40	31	40.26
Cleveland	72	47	14	29.79	33	70.21	39	54.17
Detroit	106	150	92	61.33	58	38.67	48	45.28
Humboldt County	34	33	17	51.52	16	48.48	18	52.94
Kansas City	102	31	8	25.81	23	74.19	79	77.45
Maine	67	44	22	50.00	22	50.00	45	67.16
Memphis	47	22	13	59.09	9	40.91	38	80.85
Minnesota	24	19	8	42.11	11	57.89	13	54.17
Oregon	92	79	16	20.25	63	79.75	29	31.52
Puget Sound	104	90	41	45.56	49	54.44	55	52.88
South Central PA	85	55	19	34.55	36	65.45	49	57.65
West Michigan	112	80	33	41.25	47	58.75	65	58.04
Western New York	293	103	42	40.78	61	59.22	232	79.18
Wisconsin	67	37	7	18.92	30	81.08	37	55.22
Total	1282	896	392	43.75	504	56.25	778	60.69

discussed below, the instability in MHCA membership has significant implications for measurement and analysis of MHCA organization and organizational change.

Methods Issues in Assessing Change in Alliance Attributes

Because of high turnover in MHCA membership, a central question becomes whether any observed differences in MHCA organization over time represents change in basic alliance characteristics or, alternatively, differences in perceptions of a combination new and, perhaps self-selected, ongoing members of the MHCA. To take a hypothetical example, a subset of alliance members who were positively disposed toward leadership at time 1 might have become disillusioned and dropped out of the alliance. They may have then been replaced by new members who were predisposed to give leadership positive evaluations, and when combined with retained members committed to leadership, could have resulted in high leadership effectiveness ratings at

time 2. Given that MHCAs are essentially synonymous with their membership, such dynamics invite careful measurement approaches and interpretation to understand MHCA change. Table 2, for example, indicates the mean level of leadership for the 14 alliances between the two rounds of the survey. For a number of alliances, these measures of leadership reveal little change over the two time periods. For example, the mean level of leadership for Cincinnati was 4.22 at time 1 and 4.24 at time 2. While one interpretation might emphasize little movement in MHCA structure or practice, another might suggest that a change in membership reinforced support for existing practices by weeding out members who were less supportive of those practices, and bringing in new members who are more supportive of MHCA leadership. To take another example, Maine appears to show fairly substantial differences between time 1 (M = 4.17) and time 2 (M = 4.45). Such differences may reflect change in alliance leadership processes as they evolve and mature organizationally. However, an alternative interpretation may suggest these observed differences are driven not by fundamental improvements in leadership but by differences in perceptions among the two samples of respondents in the two rounds of the survey.

Methods Issues in Assessing Construct Reliability over Time

In the context of the issues described above, an important foundation in assessing MHCA change is to first determine the stability (and thus the comparability) of MHCA organizational constructs over time. For example, do latent dimensions of leadership in alliances reflect a comparable structure at different time periods to allow investigators to infer whether leadership is becoming less effective, more effective, or remains stable? Alternatively, if the latent structure of leadership differs at different time points, what accounts for such differences and how do investigators assess change in important alliance characteristics such as leadership?

To assess how stable alliance leadership constructs were over time, a confirmatory factor analysis was conducted for the leadership construct for both time periods. Goodness-of-fit statistics confirmed that the leadership construct exhibited the same factor structure (i.e., one factor and the same items loading on these factors) across the two time periods (Table 3). The Cronbach's alphas for the leadership construct also indicate good internal consistency among the constituent items for both time periods.

These results lay the foundation for assessing how leadership effectiveness changes over time and increases the confidence that investigators have in

Table 2: Changes in the Leadership Scale—All Respondents, Panel Respondents, and Nonpanel Respondents

	All Respondents				*Panel Respondents*				*Nonpanel Respondents*			
	Round 1		*Round 2*		*Round 1*		*Round 2*		*Round 1 (Ex-Members)*		*Round 2 (New Members)*	
	n	*Mean (SD)*	*n*	*Mean (SD)*	*n*	*Mean (SD)*	*n*	*Mean (SD*	*n*	*Mean (SD)*	*n*	*Mean (SD)*
Cincinnati	52	4.22 (0.60)	59	4.24 (0.59)	26	4.26 (0.59)	26	4.15 (0.71)	26	4.17 (0.63)	33	4.33 (0.47)
Cleveland	36	4.10 (0.62)	24	4.19 (0.61)	8	4.21 (0.48)	8	4.20 (0.43)	28	4.07 (0.66)	16	4.18 (0.71)
Detroit	39	3.89 (0.73)	77	3.91 (0.68)	16	3.90 (0.68)	16	3.89 (0.74)	23	3.88 (0.78)	61	3.91 (0.67)
Humboldt	26	4.33 (0.41)	22	4.29 (0.47)	10	4.46 (0.34)	10	4.42 (0.38)	16	4.24 (0.44)	12	4.17 (0.52)
Kansas City	49	4.21 (0.63)	19	4.31 (0.58)	8	4.44 (0.45)	8	4.32 (0.73)	41	4.16 (0.66)	11	4.31 (0.47)
Maine	32	4.17 (0.58)	26	4.45 (0.47)*	8	4.52 (0.35)	8	4.63 (0.36)	24	4.05 (0.60)	18	4.36 (0.49)*
Memphis	20	4.13 (0.49)	11	4.36 (0.56)	3	4.51 (0.43)	3	4.24 (1.08)	17	4.04 (0.48)	8	4.40 (0.46)
Minnesota	17	3.93 (0.66)	15	3.83 (0.63)	8	3.71 (0.60)	8	3.57 (0.47)	9	4.12 (0.69)	7	4.16 (0.71)
Oregon	58	4.42 (0.46)	48	4.30 (0.50)	32	4.41 (0.46)	32	4.37 (0.48)	26	4.43 (0.47)	16	4.16 (0.52)*
Seattle	67	4.14 (0.51)	57	4.15 (0.49)	24	4.21 (0.38)	24	4.31 (0.37)	43	4.10 (0.58)	33	4.04 (0.54)
Western Michigan	51	3.73 (0.80)	36	3.98 (0.67)	16	4.05 (0.59)	16	4.16 (0.45)	35	3.59 (0.84)	20	3.83 (0.78)
Western New York	94	3.99 (0.80)	57	3.79 (0.68)	20	4.17 (0.54)	20	3.97 (0.62)	74	3.93 (0.87)	37	3.69 (0.70)
Wisconsin	41	3.97 (0.66)	25	4.00 (0.57)	16	4.05 (0.73)	16	3.96 (0.56)	25	3.91 (0.61)	9	4.07 (0.62)
York	52	4.00 (0.63)	35	4.15 (0.51)	19	4.08 (0.46)	19	4.07 (0.60)	33	3.96 (0.72)	16	4.25 (0.40)
Combined	634	4.09 (0.66)	511	4.10 (0.61)	214	4.20 (0.55)	214	4.16 (0.59)	420	4.03 (0.70)	297	4.06 (0.63)

*Difference between survey rounds significant at $p < 0.05$.

Table 3: Factor Loadings for Leadership Effectiveness

	Round 1	*Round 2*
The Alliance's leadership comes up with inventive ideas	0.73	0.69
The Alliance's leadership creates a climate of productive accomplishment in the Alliance	0.82	0.79
The Alliance's leadership holds regular reviews of progress on the Alliance activities	0.78	0.74
The Alliance's leadership provides helpful advice to the Alliance members	0.76	0.75
The Alliance's leadership gets things done	0.76	0.77
The Alliance's leadership is skillful in resolving conflict	0.77	0.76
The Alliance's leadership does problem solving in creative, clever ways	0.82	0.78
The Alliance's leadership has a clear vision	0.76	0.74
The Alliance's leadership facilitates efforts to develop strategic plans for the Alliance	0.72	0.68
The Alliance's leadership gets people to work productively together	0.80	0.78
The Alliance's leadership clarifies the Alliance's priorities and directions	0.77	0.74
The Alliance's leadership utilizes the skills and talents of many, not just a few	0.71	0.69
RMSEA (95% CI)	0.09 (0.08, 0.10)	0.09 (0.08, 0.10)
SRMR	0.04	0.04
CFI	0.94	0.94
TLI	0.92	0.93
Cronbach's alpha	0.94	0.94

any observed difference representing change in the initial leadership construct. However, there is no guarantee that the latent structure of the construct will be similar over time. Because of the differences in MCHA membership in the two time periods, our results could have reflected a very different factor structure in time 2 relative to time 1. Different factor structures place a burden on the investigator to determine if this reflects change in alliance membership, change in alliance structure, or simply unstable, unreliable measures.

Methods Issues in Employing Statistical Standards for Aggregating Individual Responses to Characterize Alliance-Level Attributes

In MHCAs, individual members are nested within alliances, thereby creating a multilevel structure (Javdani and Allen 2011). Although measurement of MHCA attributes is often taken at the individual respondent level, researchers are also often interested in using scores from individual respondents to

construct alliance-level measures. However, sufficient reliability and agreement among individuals who form the alliance is needed to appropriately determine if individuals' ratings of alliance characteristics can be used to represent alliance-level characteristics.

Several measures can be used to determine whether individual ratings are sufficiently homogenous to justify aggregation: r_{wg}, ICC(1), ICC(2), and the group mean reliabilities (Bliese 2000; Peterson and Castro 2006). The r_{wg} is an index of within-group agreement. For multiitem scales, as is the case in this study, this index is denoted $r_{wg(j)}$. The next two measures are both forms of the intraclass correlation coefficient (ICC). The first, ICC(1), refers to the proportion of variance in the leadership composite score that is due to the particular alliance. It is also often used as a measure of interrater reliability (i.e., the degree to which raters are substitutable) and, therefore, is recommended for use as a criterion for aggregating across raters. The second, ICC(2), refers to the proportion of variance in aggregated (alliance level) leadership values that can be attributed to true differences among MHCAs. ICC(2) is a reliability measure that provides an estimate of reliability of group means. The recommended value is 0.7 (Bliese 2000). Finally, group mean reliabilities refer to how well MHCAs can be distinguished on organizational characteristics such as leadership. The mean of the group mean reliabilities across sites is equivalent to the ICC(2) if group sizes are all equal. Again, the recommended value is 0.7 (Bliese 2006).

For each alliance and each survey round, mean leadership scores across respondents were first computed. For each alliance, the $r_{wg(j)}$, ICC(1), ICC(2), and group mean reliabilities were computed using PROC MIXED in SAS. The $r_{wg(j)}$, group mean reliabilities, and the group sizes for each of the 14 sites are given in Table 4. The $r_{wg(j)}$ is above the recommended value of 0.70 for all 14 sites, thereby providing evidence of within-group agreement and justification for aggregating the leadership measure. ICC(1) was 0.062 at time 1 and 0.068 at time 2, indicating that between 6.2 and 6.8 percent of the variability in leadership ratings is due to the alliance. The ICC(2) value in this study was 0.731 at time 1 and 0.710 at time 2, indicating that the measure has the ability to differentiate among the alliances on leadership. Group mean reliabilities ranged between 0.53 and 0.84 in the first time period and 0.40 and 0.85 in the second time period, differing from the ICC(2) values due to differences in group sizes.

In our example, the analysis supported aggregating individual responses to the alliance level. Aggregating individual responses to the alliance level, rather than focusing on the individual as the unit of analysis, can provide a

Table 4: Group Mean Reliabilities and within-Group Agreement

	Round 1			Round 2		
Site	*n*	*gmr*	$r_{wg(j)}$	*n*	*gmr*	$r_{wg(j)}$
Cincinnati	50	0.77	0.96	56	0.80	0.96
Cleveland	33	0.69	0.96	23	0.63	0.96
Detroit	36	0.70	0.94	75	0.85	0.95
Humboldt	26	0.63	0.83	21	0.61	0.97
Kansas City	41	0.73	0.96	19	0.58	0.97
Maine	30	0.66	0.96	23	0.63	0.97
Memphis	17	0.53	0.96	9	0.40	0.96
Minnesota	17	0.53	0.93	14	0.51	0.95
Oregon	56	0.79	0.98	46	0.77	0.97
Seattle	63	0.81	0.97	57	0.81	0.97
Western Michigan	42	0.74	0.93	30	0.69	0.95
Western New York	82	0.84	0.93	55	0.80	0.93
Wisconsin	36	0.70	0.95	25	0.65	0.96
York	48	0.76	0.96	34	0.71	0.97
ICC(1)		0.062			0.068	
ICC(2)		0.731			0.710	

means of mitigating some of the turnover issues that preclude a panel analysis of alliance change. However, such results are not always assured, especially in situations where there is substantial instability in alliance membership. For example, membership turnover often decreases the frequency of interaction between alliance participants, thereby presenting challenges to establishing agreement and reliability among participants on alliance attributes. This may negatively affect the ability to aggregate individual responses to the group level to represent alliance-level characteristics. Under such circumstances, it may be difficult to determine whether low levels of reliability reflect fundamental differences in opinion between participants about the alliance characteristic of interest, or simply that the addition of new members has temporarily destabilized agreement among members.

Strategies for Diagnosing and Managing MHCA Change Issues

As discussed, most approaches to assessing organizational change depend on having a panel of respondents available over time so that any observed change can be attributed to change in the organization or perceptions of its members,

rather than changes in cohorts of individuals comprising the organization. Assessing change also depends on assumptions about the reliability of organizational constructs over time, and meeting statistical standards for aggregating individual responses at multiple time points. As discussed previously, the nature of MHCAs often makes meeting these requirements difficult using standard survey methods.

These issues raise the question of what investigators of MHCA's can do if traditional methods of assessing change in MHCA organization are subject to question or alternative explanation. In the case of MHCAs, we advocate a problem specific set of diagnostic and analytic approaches that complement more traditional methods of analyzing organizational change. Because of MHCA membership instability noted earlier, these approaches may enable investigators to address alternative explanations for observed findings, or account for limitations in the use of traditional change evaluation methods. Examples of these approaches are outlined below. We present these examples for illustrative purposes and do not suggest that they are the only means by which investigators can address methodological challenges endemic to MHCA research.

Strategies for Assessing and Managing the Impact of MHCA Participant Instability

To assess the empirical implications of membership transition in alliances, three cohorts of alliance members should first be compared: (1) those who left the alliance after the first round (the ex-member subsample); (2) those who joined the alliance after the first round (the new member subsample); (3) those who stayed with the alliance for both rounds (the panel subsample). For example, if ex-members are those dissatisfied with alliance leadership, while new members tend to be more favorably disposed toward leadership, then estimates of leadership are likely to be biased upward. Therefore, separately examining changes in leadership using the entire sample, the panel subsample, the ex-member subsample, and the new member subsample may allow "proportional attribution" of overall change to turnover in membership, as well as to change in perceptions among the same members. It should be noted here that our data do not allow us to distinguish between membership turnover and nonresponse at the individual level. However, based on aggregate information, the panel attrition in our data is mostly due to the change in the sampling frame, rather than nonresponse. Moreover, nonresponse is a general

issue to almost all survey research and has been studied abundantly in the literature. Hence, we focus our discussion on membership turnover.

To illustrate, we first conducted *t*-tests on the leadership construct among the three cohorts (subsamples) defined above, using the survey data. Both cross-sectional differences and longitudinal change were examined, which allowed us to determine if there were significant changes in perceptions of leadership among the groups over the two rounds. Results are presented in Table 2. In particular, results from the *t*-tests show if changes over time are significantly different between panel members and nonpanel members. In this case, among all AF4Q alliances, only Maine shows a significant increase in perceived leadership among all respondents combined. Moreover, the increase appears to be mainly driven by nonpanel respondents. In another words, there is no significant change in perceptions among panel respondents, whereas new members from the second round have significantly higher perceptions of leadership than departing members from the first round).

Further quantification of differences across cohorts of members, while controlling for other factors observed in the data, can be achieved by a series of regression analyses using the difference in difference (DD) framework. The DD model is commonly used in estimating treatment or policy effects and can be adapted here for our purpose. In this scenario, the "treatment group" is the panel respondents. The model is specified below:

$$\mathrm{LD}_{it} = \alpha + \beta_1 P_i + \beta_2 R2_t + \beta_3 P_i * R2_t + \beta_4 X_{it} + \varepsilon_{it}$$

where "*i*" indexes respondents and "*t*" indexes the time period of the survey. *LD* is the outcome, in this case the leadership construct; *P* is the indicator for panel members; *R*2 is the indicator for round 2; and *X*'s are other observed factors (e.g., length of membership in the alliance). The two key coefficients are β_1 and β_3. Conditional on the observed factors, β_1 indicates the difference between panel and nonpanel respondents at the baseline, while β_3 reflects the difference between the longitudinal change among the panel members and the longitudinal change among the nonpanel members over the two rounds (the differential trends between the two groups).

We recommend estimating this model for the whole sample and for the subsamples of each alliance, assuming sufficient sample sizes for these analyses. Estimated coefficients and their statistical significance provide a comprehensive understanding of the impact of turnover and other observed factors (covariates) on longitudinal change in perceptions of leadership. We applied this analysis to our survey data, with a number of covariates (*X's*), including members' positions in the alliance, length of membership in the alliance, and

Table 5: The Difference in Difference Results for Comparing Changes in Leadership Scale between Panel and Nonpanel Respondents

	n (Rd 1)	*n (Rd 2)*	β_1	β_3
Cincinnati	52	59	0.037	−0.214
Cleveland	36	24	0.038	−0.115
Detroit	39	77	0.142	−0.095
Humboldt	26	22	0.01	0.283
Kansas City	49	19	0.319	−0.205
Maine	32	26	0.697***	−0.349
Memphis	20	11	0.989***	−1.144*
Minnesota	17	15	−0.43	−0.685
Oregon	58	48	0.038	0.239
Seattle	67	57	0.175	0.112
Western Michigan	51	36	0.53*	−0.524
Western New York	94	57	0.052	0.33
Wisconsin	41	25	0.355	−0.35
York	52	35	0.256	−0.384
Combined	634	511	0.203***	−0.065

*Difference between survey rounds significant at $p < 0.05$; **$p < 0.01$; and ***$p < 0.001$.

percentage of members' time devoted to alliance activities. Results are presented in Table 5. The estimates from our data showed that one of the key coefficients (β_3) was insignificant for most AF4Q sites and for the combined sample. The only exception is Memphis, which had a negative and marginally significant DD coefficient (−1.14, $p = .092$). Based on these results, we generally did not find that change in leadership perceptions among panel respondents was significantly different from change reported by nonpanel respondents, after controlling for other factors included in the model. However, based on the estimates of β_1, for three AF4Q sites (Maine, Memphis, and Western Michigan) as well as the combined sample, there are significant differences between panel and nonpanel respondents in terms of the leadership scale at the baseline. This may indicate the existence of important unobserved factors correlated with both turnover and leadership perception at baseline. However, those factors may not affect change in leadership perceptions over time, or they may affect the change similarly between the two groups, once other observed factors are controlled.

The analysis described above is proposed to gauge the scope and seriousness of the membership turnover problem, as well as to "attribute" observed overall change in alliance organizational characteristics to different potential sources. Two caveats pertain to this approach. First, the DD model can only assess the problem to the extent observed in the data. The portion of

the membership turnover problem due to "potential change" is not captured. For instance, ex-members and panel members may have similar perceptions of leadership at the time of the round 1 survey, but ex-members may have left the organization because their perception of leadership declined over time. Such changes in principle should be reflected in the longitudinal inference on leadership. However, as it is not observed in the data, neither the aggregate analysis nor the DD model may capture this. Second, if the DD model shows that a membership turnover problem exists and is driven by unobserved factors, which cannot be controlled, then without additional information, researchers may not have a good solution and need to simply acknowledge the existence and the seriousness of the problem. In such cases, results from the DD model should be incorporated into longitudinal inferences using the whole sample.

Depending on the results from the DD analyses described above, pooled cross-sectional analyses that control for time may be employed to assess change in alliance organization. Pooled cross-sectional analyses are often utilized when subjects or cases are comparable across time but when data are collected from different individuals at different points of time, or when overlap of individuals is so low as to be considered negligible (Sayrs 1989; Menard 2002). Therefore, this approach is preferred as the rate of turnover increases and the number of repeated individuals in the sample declines. It allows for an assessment of how time may affect change in the organizational attribute of interest, or how the joint effects of time and other covariates may result in different covariate effects at different time periods.

When substantive interest is in alliance-level characteristics and their relationship with other alliance phenomena, researchers must first establish whether it is appropriate to aggregate individual-level responses to the alliance level. However, as noted earlier, turnover among alliance participants may hinder such efforts and result in situations where reliability and agreement statistics do not support aggregation. Furthermore, because it is recommended that determinations about whether aggregation is justified be based on multiple statistics, there is the possibility that statistics will conflict (e.g., different aggregation statistics within a period, the same statistics across periods, or some combination of both).

One alternative for researchers under these circumstances is to revisit the theoretical rationale for why individual-level responses are being aggregated and ask whether they should expect aggregation statistics to be in agreement. For example, multilevel researchers have differentiated between compilation and composition processes (Bliese and Jex 2002). Composition

processes are based on the premise that higher level constructs are reflective of lower level phenomena and, thus, they are expected to be the same across levels. Therefore, within-group agreement is expected with composition processes while reliability is less important. In contrast, compilation processes are assumed to differ across levels such that the aggregated construct is not expected to reflect phenomenon observed at a lower level. For example, researchers may be interested in the role and influence of different stakeholder types in alliance functioning. While stakeholder measures are typically constructed from individual-level indicators that reflect stakeholder type (e.g., insurer, provider, consumer), when aggregated to the alliance level, these indicators can take on a different meaning (i.e., stakeholder diversity/concentration). Thus, there may not be any reason to expect agreement. Similarly, agreement and reliability statistics themselves may be used as variables in multilevel models, particularly when theoretical or substantive interest is in dispersion or variance among individuals (LeBreton and Senter 2008). For example, Schneider, Salvaggio, and Subirats (2002) found that the relationship between employees' perceptions of customer service climate and customer ratings of satisfaction was moderated by the strength of climate perceptions (i.e., level of within-group agreement). Both of these approaches may provide a means of examining the influence of alliance-level phenomena and their changes over time when aggregation statistics are not appropriate, or they do not support construction of alliance-level measures.

Finally, we should note that biases introduced by shifting participation in alliances can, under some conditions, be addressed though the inclusion of appropriate controls in the model. For example, including a "case mix" measure that captures, say, the demographic, organizational role, and experience mix of organizational participants at different time points may mitigate problems associated with measurement error by adjusting for the impact of participation instability. Of course, such an approach depends heavily on identifying the "right" attributes that comprise the "case mix," specifically those that collectively are associated with both dropout rates and the alliance attribute of interest.

Conclusion

The central theme of this article is that high levels of instability and turnover in MHCA membership create challenges in using survey data to study changes in key organizational characteristics of MHCAs. In addition to

describing several of these challenges, we propose analytic strategies to diagnose both the extent and source of these problems. However, these strategies are only as good as the data on which they are based. Because investigators often lack adequate information to take full advantage of these approaches, we suggest several areas of data augmentation that, where feasible, may enhance information in the data and potentially ameliorate difficulties in executing these strategies. For example, conducting "exit interviews" of departing participants to determine their reasons for departure and their perceptions of the alliance at that point. This may provide important insights on whether departing members have systematically different views of the alliance from members who stay on. Key social-demographic characteristics of respondents should be also collected, especially those likely to influence their perceptions of important alliance features. For example, an alliance whose participants are CEOs is likely to have different dynamics than those composed of lower level organizational representatives. To the extent that this composition may change over time, there may be substantive effects on the organizational role of the alliance and its performance. Such information would help investigators better control for heterogeneity across alliances when examining changes.

Given the potential of such data for improving the quality of research on MHCAs and by extension the performance and management of these organizations, the additional investment may prove cost effective. From a broader perspective, however, real solutions to the problems of quantitatively assessing change in MHCAs may be limited, and traditional, survey-based approaches may need to be supplemented with other types of data and mixed methods designs (Creswell et al. 2011; Zhang and Cresswell 2012). For example, while organizational change might be quantified using multiple waves of survey data, in depth qualitative data might be employed to account for possible explanations for any observed changes. In depth case studies of alliance, organizational change may also offer insights on these issues by addressing the complex interactions between alliance participants and between alliances and their environments, dynamics that are difficult to capture using survey methods alone.

Some may argue that the challenges confronting investigators who study alliances such as MHCAs are no different than those confronting investigators studying change in any organization. However, the susceptibility of alliances such as MHCAs to shifting levels of participation and membership makes these issues more central to the analysis of organizational change in alliances. These features of alliances are both substantively important and methodologically challenging. Rather than assuming that alliances remain stable over time,

investigators need to accept the challenge of understanding the implications of these dynamics for analysis and measurement and how they impact to alliance performance and sustainability.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was funded by the Robert Wood Johnson Foundation under the Aligning Forces for Quality Program. There are no known conflicts of interest.

Disclosures: None.

Disclaimers: None.

REFERENCES

Alexander, J., L. Hearld, and J. Mittler. 2011. "Measuring Leadership in Multi-Sector Health Alliances." *Nonprofit Management and Leadership* 21 (4): 341–362.

Alexander, J., M. Comfort, B. Weiner, and R. Bogue. 2001. "Leadership in Collaborative Community Health Partnerships." *Nonprofit Management and Leadership* 12 (2): 159–75.

Bliese, P. D. 2000. "Within-Group Agreement, Non-Independence, and Reliability: Implications for Data Aggregation and Analysis." In *Multilevel Theory, Research and Methods in Organizations*, edited by K. J. Klein, and S. W. Kozlowski, pp. 349–381. San Francisco, CA: Jossey-Bass.

——— 2006. *Multilevel Modeling in R (2.2): A Brief Introduction to R, the Multilevel Package and the nlme Package.* Washington, DC: Walter Reed Army Institute of Research.

Bliese, P., and S. Jex. 2002. "Incorporating a Mulitilevel Perspective into Occupational Stress Research: Theoretical, Methodological, and Practical Implications." *Journal of Occupational Health Psychology* 7 (3): 265.

Blumenthal, D. 2012. "Letter from the Guest Editor." *American Journal of Managed Care* 18 (6 Suppl): s95.

Creswell, J. W., A. C. Klassen, V. L. PlanoClark, and K. C. Smith. 2011. *Best Practices for Mixed Methods Research in the Health Sciences.* Bethesda, MD: National Institutes of Health.

Florin, P., R. Mitchell, J. Stevenson, and I. Klein. 2000. "Predicting Intermediate Outcomes for Prevention Coalitions: A Developmental Perspective." *Evaluation and Program Planning* 23: 341–346.

Gamm, L. D. 1998. "Advancing Community Health through Community Health Partnerships." *Journal of Healthcare Management* 43 (1): 51–66, discussion 66–57.

Huxham, C. 1996. *Creating Collaborative Advantage.* Thousand Oaks, CA: Sage Publications.

Javdani, S., and N. E. Allen. 2011. "Proximal Outcomes Matter: A Multilevel Examination of the Processes by Which Coordinating Councils Produce Change." *American Journal of Community Psychology* 47 (1–2): 12–27.

Koza, M. P., and A. Y. Lewin. 1998. "The Co-Evolution of Strategic Alliances." *Organization Science* 9 (3): 255–264.

LeBreton, J. M., and J. L. Senter. 2008. "Answers to 20 Questions about Interrater Reliability and Interrater Agreement." *Organizational Research Methods* 11 (4): 815–852.

Menard, S. 2002. *Longitudinal Analysis*. Newbury Park, CA: Sage Publications.

Okubo, D., and K. Weidman. 2000. "Engaging the Community in Core Public Health Functions." *National Civic Review* 89 (4): 309–326.

Peterson, M. F., and S. L. Castro. 2006. "Measurement Metrics at Aggregate Levels of Analysis: Implications for Organization Culture Research and the GLOBE Project." *Leadership Quarterly* 17 (5): 506–521.

Sayrs, L. W. 1989. *Pooled Time Series Analysis*. Newbury Park, CA: Sage Publications.

Schneider, B., A. N. Salvaggio, and M. Subirats. 2002. "Climate Strength: A New Direction for Climate Research." *Journal of Applied Psychology* 87 (2): 220–229.

Sink, D. 1996. "Five Obstacles to Community-Based Collaboration and Some Thoughts on Overcoming Them." In *Creating Collaborative Advantage*, edited by C. Huxham, pp 101–109. London: Sage.

Spekman, R., T. M. Forbes, L. A. Isabella, and T. C. MacAvoy. 1998. "Alliance Management: A View from the Past and a Look to the Future." *Journal of Management Studies* 35 (6): 747–772.

Swain, G. R., N. Bennett, P. Etkind, and J. Ransom. 2001. "Local Health Department and Academic Partnerships: Education beyond the Ivy Walls." *Journal of Public Health Management and Practice* 12 (1): 33–36.

Zakocs, R. C., and E. M. Edwards. 2006. "What Explains Community Coalition Effectiveness? A Review of the Literature." *American Journal of Preventive Medicine* 30: 351–361.

Zhang, W., and J. Cresswell. 2013. "The Use of "Mixing" Procedure of Mixed Methods in Health Services Research." *Medical Care*, 51 (8): e51–7. doi:10.1097/MLR.0b013e31824642fd.

Zuckerman, H. S., A. Kaluzny, and T. C. Ricketts. 1995. "Alliances in Health Care: What We Know, What We Think We Know, and What We Should Know." *Health Care Management Review* 20 (1): 54–64.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.