Evaluating systems change in a juvenile justice reform initiative

John K. Roman a,⁎,1, Jeffrey A. Butts b,2, Caterina Gouvis Roman c,3

a Urban Institute, Justice Policy Center, 2100 M Street NW, Washington DC 20037, United States
b Research and Evaluation Center, John Jay College of Criminal Justice, City University of New York, 555 West 57th Street, Suite 605, New York, NY 10019, United States
c Temple University, Department of Criminal Justice, Gladfelter Hall, 5th Floor, 1115 Polett Walk, Philadelphia, PA 19122, United States

⁎ Corresponding author. Tel.: +1 202 261 5774.
E-mail addresses: jroman@urban.org (J.K. Roman), jbutts@jjay.cuny.edu (J.A. Butts), croman@temple.edu (C.G. Roman).
1 Present address.
2 Present address. Tel.: +1 212 237 8486.
3 Present address. Tel.: +1 215 204 1025; fax: +1 215 204 3872.

A R T I C L E   I N F O
Available online 14 June 2011

Keywords:
Juvenile justice reform
Program evaluation
Reclaiming Futures
Social network analysis
Substance abuse treatment
Systems change

A B S T R A C T
Evaluating comprehensive, interagency initiatives to reform human services systems presents substantial challenges to traditional research models. Outcomes are observed at the system level rather than the individual level, and the validity of study results may be challenged on a variety of dimensions, particularly small sample sizes and measurement error. We report the results of a cross-site evaluation of the first phase of Reclaiming Futures, a five-year effort to improve services and interventions for justice-involved youth in 10 communities across the United States. We present the evaluation findings and discuss implications for similar research endeavors. The methods used in this study may be useful for other system-level evaluation efforts. A social network analysis analyzes changes in the size, density, and cohesiveness of stakeholder networks. Bivariate models test for associations between Reclaiming Futures implementation and key stakeholder perceptions of effective system change. Multivariate models are used to explore selection effects on item response. In general, we find positive correlations between successful implementation of Reclaiming Futures and indicators of effective and efficient system change.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

We report the results of a cross-site evaluation of the Robert Wood Johnson Foundation initiative, Reclaiming Futures (RF), a system change and reform strategy designed to improve the integration of drug abuse interventions with the juvenile justice process. The evaluation included two components: a social network analysis to identify changes in the pattern of agency collaboration in the first 10 communities to implement Reclaiming Futures, and a survey of stakeholders in each community to track their perceptions of the quality and coordination of youth services (For detailed discussions of these topics, see Butts & Roman, 2007; Butts, Roman, & Gitlow, 2009; Nissen, Butts, Merrigan, & Kraft, 2006; Roman, Sundquist, Butts, & Chalfin, 2010; Yahner & Butts, 2007). In Reclaiming Futures, the intent of system change was to reduce drug use problems among youth involved in the juvenile justice system through improved coordination between juvenile justice and health and human services agencies. The primary focus of the evaluation was to measure changes in communication and collaboration among partner agencies. Below, we first provide a description of the Reclaiming Futures model. We then review previous work on collaboration within a system change context. Finally, we introduce the data and methods used in the study and present the results, accompanied by a discussion of the implications of the research for future evaluations, including impediments to effective measurement of system change.

2. The Reclaiming Futures initiative

The first phase of Reclaiming Futures occurred between 2002 and 2007. The Robert Wood Johnson Foundation (RWJF) and the National Program Office (NPO) at Portland State University worked to promote more effective partnerships between juvenile justice and substance abuse treatment systems, and to foster intervention strategies for court-involved youth that would be comprehensive, proactive, evidence based, and community oriented. The RWJF and NPO launched Reclaiming Futures by awarding project grants to 10 communities in the eastern, central, and western United States. Other sites later joined the initiative, supported by additional funding sources, including the Office of Juvenile Justice and Delinquency Prevention (OJJDP), the Center for Substance Abuse Treatment...
(CSAT), and the Kate B. Reynolds Charitable Trust in Winston-Salem, North Carolina.

Reclaiming Futures is not a treatment program for individual clients, but rather a strategy for reforming and reorganizing systems at the agency, network, and community levels. It is an approach to building collaborative partnerships between juvenile justice practitioners, treatment providers, community groups, and family members with the common goal of developing comprehensive and integrated systems of care for drug-involved and court-involved youth. It is not a standardized, manual-driven process. Each Reclaiming Futures community pursues its own reform strategy, depending on its unique circumstances and characteristics.

From the very beginning of Reclaiming Futures, the Robert Wood Johnson Foundation (2001) described the purpose of the initiative as a "collaborative partnership between juvenile justice practitioners, treatment providers, and communities," an effort to develop "new systems of comprehensive care in the juvenile justice system," and a project to build "comprehensive, integrated community systems of care for substance abusing youth." The principal strategy of Reclaiming Futures was to improve communication and collaboration among practitioners, policymakers, and community members. The Foundation acted on a belief that justice-involved youth with substance abuse problems received ineffective services at least in part because communities were unable to coordinate the efforts of their juvenile justice and substance abuse treatment systems.

Participating communities were encouraged to focus their efforts on small groups of youth initially (perhaps only adjudicated youth with severe substance abuse problems), but then to expand their reach to nonadjudicated, less serious offenders, or even diverted youth. A National Program Office was established at Portland State University to manage the initiative, but the NPO directed each community involved in Reclaiming Futures to decide for itself how broad its agenda for change should be.

The Reclaiming Futures model describes an interagency, community-integrated system for responding to substance abuse problems. The diversity of implementation approaches adopted by communities was guided by the RF model (Nissen et al., 2006; also Nissen & Merrigan. "More Treatment," this issue), a relatively straightforward performance measurement model designed to help communities unite the efforts of courts, service providers, community organizations, and individual volunteers. The RF model requires juvenile justice agencies and substance abuse treatment systems to cooperate across agency boundaries to fashion better intervention approaches for youth, but it also requires them to form community-justice partnerships that would incorporate the efforts of volunteers and community groups in engaging youth in prosocial opportunities and community supports.

The RF model portrays system activities in six stages. The first two stages focus on the juvenile justice system's responsibility to identify the level and severity of a youth's involvement with drugs and alcohol using validated screening and assessment instruments. Youth identified by a screening as having drug and/or alcohol issues, according to the RF model, should be referred for a full assessment. In addition to measuring the juvenile's risks of alcohol and drug abuse or dependence, an assessment should also measure family risks, needs, and strengths.

Based on the results of the assessment, the third stage of the RF model suggests that an interdisciplinary team be assembled to develop a balanced program of interventions to address needs and risks, and to cultivate the strengths, skills, and talents of each youth and family. The interdisciplinary team should include family members and supportive members of the youth's social network, as well as representatives from juvenile justice, substance abuse treatment, mental health services, and schools. The RF model reflects a youth development approach to juvenile justice (Butts, Bazemore, & Meroe, 2010; Lerner, Dowling, & Anderson, 2003; Schwartz, 2000).

The last three stages of the RF model focus on implementing the service plan and engaging the youth and family in a meaningful way that facilitates the youth's attachments with others and increases his or her involvement in the community. Service plans should not be restricted to substance abuse treatment or even to professional services. Each service plan should encompass educational and occupational assistance, participation in prosocial activities, community, and civic engagement, and positive relationships with natural helpers or mentors.

3. Prior research

The study of collaboration has a long history of theorizing related to management and organization (Cropper, Ebers, Huxham, & Ring, 2008). There is no single definition of organizational collaboration (Alexander, 1995). Some have described the concept as any action that takes the behavior of another organization into account. Others reflect a narrower view, one limited to situations in which two or more organizations act jointly to achieve mutually beneficial outcomes. In human services, terms such as “services integration,” “collaboration,” and “cooperation” are often used interchangeably, but recognizing distinctions among these terms is useful. The term “collaboration” more closely represents what is needed for system change. Existing research defines levels of collaboration according to the extent of communication, joint planning, and decision making (Bruner, 1991; Christenson & Robinson, 1989). For instance, Bruner (1991) describes the three terms as levels on a continuum:

- **Communication** can help people do their jobs better by providing more complete information, but it does not require any joint activity. Any linkages are limited and usually isolated to one task or objective.
- **Coordination** involves joint activity, but it allows individuals to maintain their own sets of goals, expectations, and responsibilities.
- **Collaboration** requires the creation of joint goals to guide the collaborators' actions.

Studies in the field of organizational theory and public health also suggest that not only is it important to distinguish characteristics of collaboration by the types and amounts of resources shared, but also to distinguish among the types and number of agencies included in the sharing process. In this regard, characteristics of collaboration include both horizontal and vertical integration. Horizontal integration refers to communication and collaboration that occurs across sectors (e.g., substance abuse, mental health, juvenile justice) and across different areas within a particular sector (e.g., courts and police within the justice sector), but it generally involves different groups that act at the same level. Horizontal collaboration, by definition, then, moves a partnership toward system change by reducing service fragmentation. Horizontal integration is generally viewed on a continuum from communication to collaboration.

Vertical integration refers to partnerships that are communicating and collaborating regularly at different levels of government, and the integration of groups includes community-based and other nongovernmental organizations. Essentially, for vertical collaboration, nongovernmental organizations would have the ability to secure goods and services from the traditional power holders or the typically more centralized partners.

Successful implementation of integrated services systems is inherently difficult and necessitates collaboration. Thus, it is not surprising that the concepts of coordination and collaboration appear frequently in the literature and practice of human services. Researchers find that, if properly implemented, coordination itself can improve service delivery (Agranoff, 1991; Gans & Horton, 1975; Harris, Crilly, Strole, & Ellett, 1999; Hassett & Austin, 1997; Lourie, 2003; Miller, Scott, Stage, & Birkholtz, 1995; Stroul, 2003; U.S. General Accounting Office, 1992; Waldfogel, 1997). Improvement in service delivery alone, however, is not synonymous with system change.
System change is a goal broadly encompassing a long-term outcome where agencies and organizations involved in a partnership collaborate to achieve a set of mutually agreed upon best practices as part of a normal routine. System change is an evolving, iterative process that can take years to achieve. Collaboration may occur more quickly.

The Reclaiming Futures model suggests that system change is achieved through collaboration focused on services integration. Services integration becomes a priority to prevent clients from falling through the cracks between service providers. In an effort to create safety nets, comprehensive services integration links clients with existing services using such methods as colocation of the providers and case managers that are assigned to a particular client. System integration often necessitates a structural reconfiguration and reevaluation of program mission by all participating organizations. System integration efforts seek to streamline the way organizations plan, fund, collect, and share data on programs to eliminate conflicting program requirements. The result of the partnership should be a seamless system of mutually reinforcing interventions.

Although the theoretical literature on organizations and partnerships abounds with examples of what system change should look like, the empirical literature is scarce. The empirical literature that does exist focuses mostly on the fields of public health and child development, and does not articulate a set of strategies for justice-based partnerships. Generalization across fields or substantive domains is difficult because existing system change models are organized around different purposes, structured and managed in different ways, and may include different partners. Additionally, virtually all successful integrated service models are local and often inspired by local leadership. Few studies measure and document collaborative strategies in a systematic fashion that would be useful to practitioners and researchers implementing and studying system change in varying contexts (Roman, Jenkins, & Wolff, 2006).

For instance, many efforts at system reform in the human services build upon the lessons of the Child and Adolescent Service System Program (CAASP). Begun in 1984 by the Center for Mental Health Services in the U.S. Department of Health and Human Services, CAASP was designed to integrate child and adolescent services, including child welfare, juvenile justice, special education, health, and mental health services (Lourie, 2003). The results of CAASP, however, did not produce a blueprint for system change in particular sectors, including juvenile justice.

Other important lessons in system change can be found in the evaluation of the Fort Bragg Demonstration Project, a five-year initiative that provided funds to improve juvenile mental health and substance abuse services (Bickman, 1996). Researchers found that the $80 million project was implemented successfully. Access to services improved, and treatment settings became less restrictive. However, the services offered through the demonstration project cost more than traditional services, and clinical outcomes for clients in the demonstration sites were no different than outcomes in the comparison sites. Thus, key lessons emerge: (1) Successful implementation of complex system change initiatives does not always signify that long-term client outcomes will be achieved. (2) Furthermore, even success in achieving short-term outcomes, such as improved access to services and an improved suite of treatment services, does not imply that long-term outcomes will be achieved. (3) Finally, if successful implementation of a complex initiative comes at such a cost (i.e., $80 million), would the initiative have cost even more had it been successful with regard to long-term client outcomes?

In general, studies find that collaboratives in which power and resources are shared across organizations, and in which multiple links exist between pairs of agencies within the network, tend to be more successful than partnerships lacking such strong connections (Provan, Nakama, Veazie, Teufel-Shone, & Huddleston, 2003; Roman et al., 2006). Successful partnerships are those where all (or most) organizations communicate with one another and where each agency can leverage the resources it needs to achieve its goals; and the most effective partnerships appear to be those in which community-based organizations form strong relationships with government agencies (Abbot, Jordan, & Murtaza, 1993; Burns & Spilka, 1997; Hodges, Nesman, & Hernandez, 1999). More specifically, research has suggested that partnerships that achieve both horizontal and vertical integration will be the most successful with regard to system integration (Roman et al., 2006). These closely linked collaborative processes, over time, help build shared values, purpose, trust, and leadership—elements of successful partnerships in their own right.

Although the research literature is rife with best practices for collaboration that describe common obstacles to system change, far fewer studies have described the measurement of system change factors such as horizontal and vertical integration. For partnerships that expressly seek system change, assessing horizontal integration is particularly important because organizations may be present for meetings but may not show the commitment and level of cooperation that takes them beyond their own goals and objectives.

Given the dearth of evaluation literature focused on measuring system change in human service systems, in recent years, a growing number of social science researchers have begun to borrow principles of social network analysis to assess partnerships and system change (see for example: Feinberg, Riggs, & Greenberg, 2005; Hefflinger, 1996; Morrissey, 1992, 1997; Provan & Milward, 1995). Social network analysis refers both to a perspective for examining social relations and to a methodological technique for analyzing those relations. Social network analysis views relations between actors as patterns and, eventually, structures that shape the behavior of individuals or groups (Marsden, 1980). Social network data describe the contacts, ties, and attachments that one individual or group has to another. By examining network data, researchers can reconstruct the patterns of interaction and social structures that influence behavior.

In fields such as sociology, anthropology, and social psychology, social network analysis is used to analyze the strength and quality of social ties between people or organizations in terms of structures, roles, and patterns of communication. Despite its versatility, researchers are just beginning to use social network analysis to assess the effects of interorganizational collaborations in justice settings. The few studies in this field include those by George Kelling et al. (1997), who used network analysis to describe and compare communication networks in sites implementing the federally sponsored Comprehensive Communities program. Sridharan and Gillespie (2004) used network analysis techniques to understand collaborative networks across sites participating in a different community change initiative. More recently, a Pennsylvania-based study used social network analysis to assess the relationship between network dynamics and community readiness to engage in the organizational change processes required by the Communities That Care program (Feinberg et al., 2005).

In the present paper, we use network analysis as an exploratory tool to confirm findings from the stakeholder survey, assess factors of collaboration that are more rigorously captured using the quantifiable principles of social network analysis, and examine change over time. Collaborative factors such as horizontal and vertical integration can be operationalized in social network techniques using matching network variables such as proximity, cohesion, and power equity.

4. Data and methods

The primary source of data for the evaluation of Reclaiming Futures was a repeated measures survey of 20 to 50 stakeholders in each of the 10 Reclaiming Futures sites. The survey included two key components:

- A set of 58 questions about the effectiveness of the local youth services system’s capacity to design and deliver coordinated service plans for drug-involved youthful offenders
• A second set of questions about the key stakeholders in each community, including key officials, allied professionals, volunteers, and other community actors, that asked how often they interacted with one another, (1) whether the interactions were helpful to each respondent in doing his or her job, and (2) whether the respondent believed the same interactions were helpful to others in doing their jobs.

Two analyses were performed for the evaluation of Reclaiming Futures. First, using data from the stakeholder survey, a series of statistical tests assessed whether the initiative was associated with improved respondent perceptions of system effectiveness. Second, using the data on social relationships from the stakeholder survey, a social network analysis tested whether changes in network density, cohesion, and proximity suggested that overall network strength increased in Reclaiming Futures communities.

4.1. Stakeholder surveys

The survey items used were constructed specifically for the study. After reviewing the literature in human services management, healthcare, substance abuse treatment, and children services, the evaluation team determined that no existing instruments would serve our purpose. A number of researchers and writers, however, were important influences on our thinking and our approach to measurement (Damanpour, 1991; Glisson, 1989, 2002; Lehman, Greener, & Simpson, 2002). Prior to the first wave of data collection, pilot testing was conducted by distributing an early version of the survey to a group of 10 initiative partners, who completed the study and returned comments.

Prior to the second wave, individual questions were aggregated within parent constructs, and construct analyses were performed. The goal of this analysis was to determine how well survey items measured each construct, whether other constructs emerged, and whether any constructs overlapped. We also wanted to identify any items that could be omitted from the survey. First, we examined the correlation among all items. We expected items to be highly correlated, because we believed they measure the same construct. But we also wanted to identify items with such high correlations that they likely contributed the same information—meaning that one could be omitted to shorten the survey. Next, we performed a confirmatory factor analysis (principal axis factoring) to determine whether variables loaded on an underlying construct. We then conducted a reliability analysis to determine how strongly items held together. Reliability coefficients are affected by how strongly items correlate with one another and by the number of items included (more items equals higher reliability). Therefore, it is possible to obtain a high reliability coefficient for a large number of items even when items are not strongly correlated. Finally, we conducted an exploratory factor analysis (principal components analysis) of all variables to observe the number of underlying constructs and the variables that loaded highly on them.³

As a result of these analyses, about 20 items were dropped from the original survey. The composition of a number of constructs was changed as additional constructs emerged, and constructs were renamed to better describe them. The reliability of the scales was reestimated following each subsequent survey wave, through the fourth wave. The reliability coefficients for each construct wave were above 0.7, which is the standard convention for establishing strong reliability.

The Reclaiming Futures stakeholder surveys were administered six times between 2003 and 2006. The evaluation team selected the stakeholder respondents in each community after consulting with the project director from that community. Researchers instructed each project director to identify the most informed 20 to 40 local practitioners who would have the greatest knowledge of the youth services system in that community. In each community, no more than two individuals worked for Reclaiming Futures. The most common Reclaiming Futures configuration was a project coordinator (often less than full-time) and an office administrator or administrative assistant. Beyond the one or two people paid by the initiative, each site had five “fellows,” these were unpaid advisors from each of the key sectors who met periodically with the project coordinator. Thus, the majority of respondents had no direct responsibility for the Reclaiming Futures initiative, but nonetheless were part of the network of individuals who could possibly serve Reclaiming Futures youth. Before each survey round, project directors reviewed and reaffirmed the list of respondents. The list could be modified to reflect job changes, retirements, new hires, new organizations, and simple omissions in a prior list. Respondents typically included judges and other juvenile justice leaders, drug treatment providers, attorneys practicing in juvenile court, community volunteers, and members of various youth-serving and faith-based organizations. When completed, each list was considered to be the finite population of ideal informants in a community rather than a sample of all possible informants.

Approximately 350 respondents were contacted for each survey wave, or about 35 people in each community. An average of 235 people responded to each wave. The total response rate across all 10 communities averaged 70%, with the lowest response rate (63%) in the first wave (December 2003) and the highest (73%) in the third wave (December 2004). The rate of respondent turnover was slight but steady. A total of 456 unique respondents completed at least one survey across six survey administrations. About half the respondents (47%) answered three or more surveys. Only 67 respondents (13%) answered all six surveys. This was not unexpected and reflects the changing composition of stakeholders over time (the response rate by site by wave can be found in Butts and Roman, 2007).⁶

The survey included 58 questions compiled into 13 multiquestions designed to measure aspects of system change. For example, the study measured the overall effectiveness of treatment by asking respondents to agree or disagree (on an interval scale) with several statements referring to the “past three months”:

• The substance abuse treatment needs of youth in my community were adequately met.
• The mental health needs of youth in my community were adequately met.
• Graduated sanctions were used effectively to support treatment goals for youth.
• Youth-serving agencies in my community generally did a good job serving youth.
• Youth-serving agencies in my community were usually able to provide youth with the range of services they needed.

Changes in the magnitude and direction of these scales were critical outcome measures for the Reclaiming Futures study. The 13 indices of systemic change included four indices related to the general concept of Administration, including Access to Services (the ease of client access to services and treatment), Data Sharing (the integration and sharing of information systems among agencies), Systems Integration (interagency coordination of policies and procedures), and Resource Management (organization, leverage of staff, and funding). Three indices measured by the survey were related to the concept of Collaboration, including Client Information (agencies

⁶ Attrition in this type of analysis cannot be evaluated using traditional diagnostics. The goal of the survey was to interview individuals who held positions that were affiliated with (but not necessarily part of) Reclaiming Futures: less than half of respondents had a direct relationship with Reclaiming Futures, and no more than 20% had a formal connection to the initiative. Thus, as individuals changed positions during the course of the evaluation, the targeted respondents changed as well (see Section 4.1.1., below).

³ A complete description of these analyses is available from the author.
sharing client information to support treatment planning), Partner Involvement (the extent of interaction among RF partner agencies), and Agency Collaboration (the quality of interagency relationships in the youth services field). Finally, six indices addressed the Quality of substance abuse treatment for youthful offenders, including Targeted Treatment (the availability of treatments appropriate for specific client groups), Treatment Effectiveness (the scope and impact of treatment services), Alcohol or Drug Assessment (the availability and use of effective screening and assessment tools), Family Involvement (the role of family members in designing and delivering services for youth), Cultural Integration (cultural competence and responsiveness), and Prosocial Activities (the use and availability of prosocial activities for youth as a part of substance abuse interventions).

Index scores for each measure were calculated as:

$$\sum_{i=1}^{n} \sum_{j=1}^{k} \left( R_{ij} \right)/nk$$

where the mean response is the sum of all responses \( R \) for each of \( l=1...n \) participants summed across all the \( j=1...k \) items in that scale and divided by the product of the number of respondents and the number of items in the scale. Responses on the 5-point scales were scored from \(-10\) (strongly negative), \(-5\) (somewhat negative), \(0\) (neutral), \(+5\) (somewhat positive), to \(+10\) (strongly positive). Some questions were originally worded negatively, but all answers were recoded so that high scores indicated positive opinions. A Wilcoxon Matched-Pairs Signed-Ranks Test was conducted on the changes across successive surveys (e.g., comparing results from wave 1 to wave 2), across all sites and also for within-site comparisons.

4.1.1. Tests for item response bias

We test for the possible effects of response bias, which would threaten the validity of our findings if respondents who are more proximate to the program report better outcomes, the results will be biased. In this study, however, the assumption is strengthened because the main effect is the change in ratings rather than a cross-sectional comparison. Still, to test for response bias, we identified the four individuals designated by each Reclaiming Futures community to be the local members of four national fellowships organized by the National Program Office. The fellowship groups included a judicial fellowship, a treatment fellowship, a justice/probation fellowship, and a community fellowship. Along with the project director, the members of the four fellowship groups served as a local leadership committee in each community and were thus more connected to the initiative than the average respondent.

To test for response bias, ordinary least squares regressions estimated the effect of leadership status on each of the 13 change indices measured by the surveys. The mean scores of the 13 indices were regressed on four variables independently: the survey wave number (because scores in general tended to increase across sites and across individual respondents); the respondents leadership status (i.e., whether a respondent participated in a fellowship group); a site dummy; and an interaction term that captured the combined effect of leadership status with a respondent’s participation in each administration of the survey. For each of the 13 change indices, the regression model was estimated separately for leaders and nonleaders, and only for respondents who participated in at least three of the six surveys.

To portray the results of the analysis, the evaluation team plotted the predicted index scores suggested by each model for both respondent groups (see Butts & Roman, 2007, for a more detailed discussion of this analysis). By plotting scores in the first survey (the Y-intercept) and then adding the predicted regression coefficient for each subsequent survey to the previous score, the analysis produced a direct comparison of leaders and nonleaders.

The analysis confirms that respondents who were more connected to Reclaiming Futures had generally (though not always) more positive perceptions of the changes resulting from Reclaiming Futures. The meaning, however, is ambiguous. Those chosen to be Reclaiming Futures fellows tended to be more experienced in substance abuse and juvenile justice programs and policies. Their positive perceptions may indicate a bias toward reporting positive changes, or it may indicate that they were better able to identify true changes. It might also be the case that the proximity of fellows to the initiative allowed them to observe positive effects that were obscured for distal observers.

4.2. Social network analysis

Social network analysis is a useful methodological tool for assessing agency collaboratives and how they change over time. Researchers have modeled organizational collaboration and coordination in varying ways using network measures (Bolland & Wilson, 1994; Hefflinger, 1996; Provan & Sebastian, 1998). The potential for network-based analyses to support and inform evaluation of collaborative partnership initiatives in criminal and juvenile justice is clear and growing (Ferguson, 2002; Kelling et al., 1997; Sridharan & Gillespie, 2004). For the Reclaiming Futures evaluation, two items were appended to the stakeholder surveys to facilitate social network analysis: “In the past 3 months, how frequently did you interact with X for work-related reasons (including meetings, e-mail, and phone calls)?” and “How helpful is this person to you in doing your job?” Although network data were collected from individual respondents, their answers were aggregated into organizational groups for reporting purposes. Thus, we make the assumption that individual responses, when aggregated within organizations, are representative of each organization’s communication and collaboration patterns.

The question items allowed the study to assess whether a key objective of Reclaiming Futures was achieved—that is, enhanced communication and collaboration among the agencies that share responsibility for planning and implementing interventions for drug-involved youth. Social network analysis was used to describe the size and structure of networks, the positions and roles played by network members, the communication and interaction patterns among members, and whether the networks evolved over time. Although we were particularly interested in capturing network properties of collaboration that would represent horizontal and vertical collaboration, to limit respondent burden, we only collected information on interactions between individuals at the organizational level. For instance, we did not ask questions about the type of interaction, such as information or resources shared within the interaction. As such, network properties can assess integration, and to some extent horizontal integration, but not vertical integration.

Network data were drawn from three of the six Reclaiming Futures surveys—June 2004, June 2005, and June 2006. Focusing on these surveys simplified the presentation of results and allowed the analysis to examine network dynamics over an extended period of time. Results of the social network analysis are reported for only 8 of the 10 Reclaiming Futures communities. In two communities, responses to surveys were not detailed enough to support network analysis, in that they were small communities with a limited number of organizations involved in the initiative.

The social network analysis generated several statistics that assessed distinct network properties relating to the presence or absence of interactions among network members (Scott, 2000). These properties indicate the size, density, and cohesiveness of a network, the power equity and instrumental equity among network members, and the relational proximity of agencies measured as the average number of ties required to establish a connection path between agency members. Table 1 provides definitions for all network measures. Network size refers to the total number of agencies in the
network regardless of whether each agency was represented by one individual respondent or several respondents. Networks ranged in size from 11 to 23 agencies. Network size was partly a function of the individual response rate in each survey (i.e., as the number of respondents increases, the number of agencies they represented was likely to increase as well).

Proximity refers to the average minimum distance between the various agencies in the network, as measured by the number of ties (interactions) required to connect any two agencies. A proximity score of 1 means that two agencies communicated directly with each other; in other words, there was a direct relationship between them. If a justice agency had a proximity of 2 in relation to a treatment agency, for example, this would mean that the shortest path of communication between the two agencies was through, or mediated by a third agency. Proximity is the only network statistic where smaller figures indicated greater performance. Note that proximity captures interaction, but not necessarily collaboration.

Density and cohesion are similar to proximity, but density refers to the proportion of all possible interactions in a network that are present (regardless of direction). Cohesion refers to reciprocal ties, or interactions that are reported by both parties, and it is measured as the proportion of all possible reciprocal interactions present. For both density and cohesion, larger percentages indicate more extensive interaction among the agencies in a network. Both density and cohesion are important constructs that can be construed as proxy measures for horizontal integration; as density and cohesion increase across all agencies, the greater the likelihood that collaborative resource sharing and full articulation of common goals across systems has occurred.

Power and influence, as described previously, refer to incoming and outgoing interactions, respectively. Agencies with many incoming ties are said to be in a position of power, whereas agencies with many outgoing ties are said to have influence over others in a network. In this analysis, we follow Kwiat, Valente, and Celentano (2001), who perform a network analysis on HIV/AIDS services providers. They describe incoming interactions (power) as the “in-degree” score of an agency or an organization, which is “the number of organizations in the network that reported referring adult HIV-positive clients to it for direct services” (p. 475). Outgoing interactions (balance), on the other hand, is described as the “out-degree” score of an organization or agency, which is “the number of organizations in the network to which it reported referring adult HIV-positive clients for direct services” (p. 476).

In our evaluation, we have asked respondents about helpfulness of interactions rather than direct referral, and thus rename in-degree scores as “power equity” and out-degree scores as instrumental equity.

If a network scores low on power equity, relatively few agencies would be responsible for all of the helpful interactions reported across all network members. A high power equity scores indicates that helpful interactions are spread more evenly. A network with a low instrumental equity score is one in which very few agencies report having helpful interactions with other agencies, whereas a high instrumental equity score indicates that helpful interactions are reported by many agencies. For both power and instrumental equity, higher percentages indicate greater equality in the distribution of ties. Similarly, both constructs taken together offer a partial picture of vertical integration where interactions are reported as helpful by most or all agencies. Because organizations with the networks represent various levels and we aggregated respondents within organizations, we make the assumption that greater power and instrumental equity represent an aspect of vertical integration. As discussed earlier, past research indicates that partnership success is more likely to be achieved if the traditional non-power holders have access to information and resources normally found at higher levels.

The analysis also generated sociograms for each community, shown in Fig. 1. Each symbol in a sociogram identified a particular agency, with the shape and color of the symbol indicating the type of agency. The location of each agency within the sociogram (in the middle or the periphery) suggests the extent of the agency’s relationships with others (Borgatti, Everett, & Freeman, 2002). Sociograms were constructed using a spring-embedding algorithm that situated each agency according to the strength and quality of its relationships with other agencies in the network. Agencies in the center of a sociogram tend to be those connected to the largest number of other network members, whereas agencies on the periphery are usually those least connected to others in the network. By convention, each arrow in a sociogram indicates a helpful tie between two agencies. Arrows point toward the agency described as helpful. A powerful agency is one with many incoming arrows, indicating that many other respondents report being helped by their interactions with the agency. A highly instrumental agency, on the other hand, is one with many outgoing arrows, as each suggests a relationship that helps that agency meet its own goals.

The network statistics associated with each sociogram provide numeric comparisons of the strength and performance of the networks. One limitation of the analysis is that direct comparisons of network statistics across networks of different size can be misleading. Small networks have fewer possible ties, and relationships between members may be more easily established and maintained. It would be unfair to expect large networks to have the same density and cohesion among partner agencies as reported by small networks. To accommodate this limitation, network statistics were analyzed only by comparing them to the average scores that would be expected for a network of equal size. To compute averages, network statistics were regressed on network size independently within waves for all six sequential surveys across each of the eight different communities. The analyses used ordinary least squares regression to measure the relationship between network size (the independent variable) and each network statistic (the dependent variables) within each wave.

The results of these analyses were used to compute average values (based on the observed trend line) that would be expected for networks of a particular size. These predicted values were then used to judge where and when networks performed at or better than average during a particular survey administration. Each statistic measured a characteristic of the network at three points in time, and every statistic was compared with the average value expected for a network of equal size. Network scores were divided into three groups: (1) strong scores (two asterisks, **) were 10% or more above the expected average; (2) positive scores (one asterisk, *) were equal to or greater than the expected average but less than 10% greater; and (3) unmarked scores fell below the expected average.

There are three notable challenges to this research. First, we did not have a census of all organizations in the Reclaiming Futures

---

Table 1

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Definition</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Number of member agencies in the network.</td>
<td>N</td>
</tr>
<tr>
<td>Proximity</td>
<td>Average minimum number of ties needed to establish a relationship between any one member agency and another.</td>
<td>1 to (N – 1)</td>
</tr>
<tr>
<td>Density</td>
<td>Network ties as a percentage of all possible ties.</td>
<td>0 to 100%</td>
</tr>
<tr>
<td>Cohesion</td>
<td>Reciprocal network ties as a percentage of all possible reciprocal ties.</td>
<td>0 to 100%</td>
</tr>
<tr>
<td>Power equity</td>
<td>Equity in the distribution of incoming ties, where 0% indicates that only one agency is responsible for all of the helpful interactions reported by network members, and 100% indicates that credit for helpful interactions is spread equally among all agencies.</td>
<td>0 to 100%</td>
</tr>
<tr>
<td>Instrumental Equity</td>
<td>Equity in the distribution of outgoing ties, where 0% indicates that only one agency reports having helpful interactions with other agencies, and 100% indicates that helpful interactions are reported equally by all agencies.</td>
<td>0 to 100%</td>
</tr>
</tbody>
</table>

---

Network statistics as described previously, refer to incoming and outgoing interactions, respectively. Agencies with many incoming ties are said to be in a position of power, whereas agencies with many outgoing ties are said to have influence over others in a network. In this analysis, we follow Kwiat, Valente, and Celentano (2001), who perform a network analysis on HIV/AIDS services providers. They describe incoming interactions (power) as the “in-degree” score of an agency or an organization, which is “the number of organizations in the network that reported referring adult HIV-positive clients to it for direct services” (p. 475). Outgoing interactions (balance), on the other hand, is described as the “out-degree” score of an organization or agency, which is “the number of organizations in the network to which it reported referring adult HIV-positive clients for direct services” (p. 476).

In our evaluation, we have asked respondents about helpfulness of interactions rather than direct referral, and thus rename in-degree scores as “power equity” and out-degree scores as instrumental equity.

If a network scores low on power equity, relatively few agencies would be responsible for all of the helpful interactions reported across all network members. A high power equity score indicates that helpful interactions are spread more evenly. A network with a low instrumental equity score is one in which very few agencies report having helpful interactions with other agencies, whereas a high instrumental equity score indicates that helpful interactions are reported by many agencies. For both power and instrumental equity, higher percentages indicate greater equality in the distribution of ties. Similarly, both constructs taken together offer a partial picture of vertical integration where interactions are reported as helpful by most or all agencies. Because organizations with the networks represent various levels and we aggregated respondents within organizations, we make the assumption that greater power and instrumental equity represent an aspect of vertical integration. As discussed earlier, past research indicates that partnership success is more likely to be achieved if the traditional non-power holders have access to information and resources normally found at higher levels.

The analysis also generated sociograms for each community, shown in Fig. 1. Each symbol in a sociogram identified a particular agency, with the shape and color of the symbol indicating the type of agency. The location of each agency within the sociogram (in the middle or the periphery) suggests the extent of the agency’s relationships with others (Borgatti, Everett, & Freeman, 2002). Sociograms were constructed using a spring-embedding algorithm that situated each agency according to the strength and quality of its relationships with other agencies in the network. Agencies in the center of a sociogram tend to be those connected to the largest number of other network members, whereas agencies on the periphery are usually those least connected to others in the network. By convention, each arrow in a sociogram indicates a helpful tie between two agencies. Arrows point toward the agency described as helpful. A powerful agency is one with many incoming arrows, indicating that many other respondents report being helped by their interactions with the agency. A highly instrumental agency, on the other hand, is one with many outgoing arrows, as each suggests a relationship that helps that agency meet its own goals.

The network statistics associated with each sociogram provide numeric comparisons of the strength and performance of the networks. One limitation of the analysis is that direct comparisons of network statistics across networks of different size can be misleading. Small networks have fewer possible ties, and relationships between members may be more easily established and maintained. It would be unfair to expect large networks to have the same density and cohesion among partner agencies as reported by small networks. To accommodate this limitation, network statistics were analyzed only by comparing them to the average scores that would be expected for a network of equal size. To compute averages, network statistics were regressed on network size independently within waves for all six sequential surveys across each of the eight different communities. The analyses used ordinary least squares regression to measure the relationship between network size (the independent variable) and each network statistic (the dependent variables) within each wave.

The results of these analyses were used to compute average values (based on the observed trend line) that would be expected for networks of a particular size. These predicted values were then used to judge where and when networks performed at or better than average during a particular survey administration. Each statistic measured a characteristic of the network at three points in time, and every statistic was compared with the average value expected for a network of equal size. Network scores were divided into three groups: (1) strong scores (two asterisks, **) were 10% or more above the expected average; (2) positive scores (one asterisk, *) were equal to or greater than the expected average but less than 10% greater; and (3) unmarked scores fell below the expected average.

There are three notable challenges to this research. First, we did not have a census of all organizations in the Reclaiming Futures
network (due to limited nonresponse). Second, we were bound to preserve the anonymity of respondents, and thus it was necessary for reporting purposes to group individual respondents into sectors to preserve anonymity (although we stress that all analysis was done at the organizational level). Third, the anonymity requirement meant that we were unable to observe participant status within an organization. Responses were aggregated into organizations and reported by sector (e.g., justice, judicial, treatment, project directors, and representatives of community organizations). Each sector includes one or more organizations within that sector, and each organization includes one or more individual respondents. The individual responses of survey participants from the same organization were averaged within organizations when multiple respondents reported on an external organization. We assert, but did not test, that

the strength of relationships at the network level can be inferred from organization relationships reported by individual respondents.

5. Results

The study results suggest that the Reclaiming Futures strategy has the potential to affect agency processes and performance. Evaluation findings are presented first for the social network analysis and then for the analysis of changes in stakeholder perceptions.

5.1. Social networks

The social network analysis suggested that Reclaiming Futures communities generally improved their network performance during

\[ \text{Network Size} \]
\[ \text{Proximity} \]
\[ \text{Density} \]
\[ \text{Cohesion} \]
\[ \text{Power Equity} \]
\[ \text{Instrumental Equity} \]

\[ \text{Network Size} \]
\[ \text{Proximity} \]
\[ \text{Density} \]
\[ \text{Cohesion} \]
\[ \text{Power Equity} \]
\[ \text{Instrumental Equity} \]

\[ \text{Note: Direct comparisons of network statistics across networks of different size, even within the same community, are inappropriate. Each statistic should be compared to the average value (in parentheses) for a network of the same size.} \]

\[ \text{1. Average expected value for a network of the same size.} \]

\[ ** = \text{Strong Performance (10\% or more above average)} \]

\[ * = \text{Positive Performance (from average to less than 10\% above average)} \]

\[ \text{LEGEND} \]

\[ \text{Project (RF)} \]
\[ \text{Court} \]
\[ \text{Justice} \]
\[ \text{Treatment} \]
\[ \text{Community} \]

Fig. 1. Sociograms and network statistics for Reclaiming Futures communities, 2004–2006.
the initiative. One simple measure of whether social networks improved is the count of positive network statistics over time. When aggregated, the communities generated 19 positive statistics in 2004, 24 positive statistics in 2005, and 26 positive statistics in 2006, an increase of 37% (Fig. 1).

The strong role played by nonjustice partners in each community was clear in the sociograms, and this was a central focus of Reclaiming Futures. As expected, however, nonjustice agencies were often clustered at the outer edges of the sociograms, suggesting that nonjustice agencies were still less central and less powerful in the juvenile justice system than were courts and probation agencies. Some agencies appeared to be relatively isolated within their networks, and these agencies were usually either a community partner or a treatment agency.

The top-performing networks in the Reclaiming Futures initiative were Seattle and Anchorage. The majority of their network statistics were at or above average across all three years of network measures (Table 2). With regard to density (the proportion of helpful interactions in any direction) and cohesion (the percentage of reciprocal interactions), the top-performing communities overall were Anchorage, Seattle, Santa Cruz, and Dayton. New Hampshire had cohesion levels at or above average during two of the three surveys. Marquette’s score on power equity was at or above average in all three surveys, and New Hampshire scored at or above average on instrumental equity in all three surveys. Considering proximity scores alone, Seattle, Santa Cruz, New Hampshire, Southeastern Kentucky, and Portland consistently performed at or above expectations given their network size.

The results of the social network analysis are suggestive only, but the evaluation revealed many positive indicators of network performance, and the indicators increased over time in several communities. Comparisons of Reclaiming Futures networks with the average expected values for similar-sized networks can be used to indicate strong communities. For example, some of the strongest communities in 2004 increased their performance further by 2006 (e.g., Santa Cruz), whereas other communities increased their network performance only slightly during the initiative (e.g., Kentucky, Marquette, Dayton, and Portland), and one community that began with relatively weak network statistics in 2004 achieved a rating of strong for most indicators in 2006 (New Hampshire).

However, it is also important to point out that, when changes in network statistics for each specific network property are assessed, improvement by network property appears to vary greatly within each community. For instance, a site may show increased cohesion between 2004 and 2006, but reduced power equity. Fig. 1 shows that only two sites (Southeastern Kentucky and Seattle) had more positive
network properties (or remained mostly the same on some and increased on others) for all network variables in 2006 compared to 2004. Although Southeastern Kentucky improved on all network properties, the site did not meet the average expected value on four of the five properties.

5.2. Stakeholder perceptions

The analysis of stakeholder surveys suggests that the Reclaiming Futures initiative generated many positive changes (Table 3). Across all demonstration sites, 12 of the 13 indices of systemic change showed significant improvements between the first and sixth survey. The largest improvements in stakeholder perceptions were observed in the Treatment Effectiveness index (growing from 0.3 to 2.8 between 2003 and 2006), the Alcohol and Other Drug (AOD) Assessment index (from 2.6 to 4.7), the index measuring the use of Pro-Social Activities for Youth (from −0.2 to 1.8), and the Data Sharing index (from 0.3 to 2.0). The one index that did not improve overall, Partner Involvement, was the highest ranked index in all six surveys, so there was little room for improvement. Availability of Client Information and Resource Management were other areas in which respondents rated their systems highly from the very beginning of the Reclaiming Futures initiative.

Since the scores are truncated between −10 and +10, improvements are constrained by the initial score, which creates a potential for bias. If a community index score was high in the first survey, there is little room for improvement in subsequent surveys. Conversely, gains are easier to achieve in a relatively low-scoring index. To adjust for this bias, the evaluation created an alternative measure for assessing general trends in the indices of system change. The Percent of Possible Improvement Realized (Table 4) measure examines whether an index realized its potential for improvement. The measure was calculated by calculating the total difference from Score 1 to Score 6, and dividing by the maximum possible improvement between Score 1 and +10.

Looking at the overall change in Reclaiming Futures communities using this alternative measure alters the results, but only slightly. The Treatment Effectiveness and AOD Assessment indices remained the two strongest performers in terms of the total change perceived by stakeholders between the first and sixth surveys. The AOD Assessment, Treatment Effectiveness, Client Information, and Family Involvement indices all realized more than 20% of total possible improvement based upon their scores in the first survey. The Client Information index moved up to third largest overall change, and the Pro-Social Activities index slipped from third-largest to sixth-largest change. The Partner Involvement index remained the only scale that
did not achieve a statistically significant positive change between the first and sixth surveys.

To judge overall change in stakeholder perceptions, the analysis considered each of the changes that occurred across all 13 indices between each of the six surveys (i.e., for a total of 65 intersurvey changes). If Reclaiming Futures had no actual effect on stakeholder perceptions in the demonstration communities, we would expect to see these intersurvey changes be half positive and half negative. On the other hand, if Reclaiming Futures had a positive effect, more of the intersurvey comparisons would be positive than negative. If the ratio of positive to negative score changes was greater than would be expected by chance, this provides additional evidence that Reclaiming Futures had an effect.

Table 2
Butts and Roman (2007, p.15).

<table>
<thead>
<tr>
<th>Community</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchorage, Alaska</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Santa Cruz, California</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Southeastern Kentucky</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Marquette, Michigan</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Dayton, Ohio</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Portland, Oregon</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Seattle, Washington</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Note: Strong network statistics are those where the actual value is above the average for a network of the same size. Five statistics are analyzed each year: proximity, density, cohesion, power equity, and instrumental equity.

Table 3
Change in mean perception scores in all Reclaiming Futures communities between December 2003 and June 2006, ranked by total amount of change.
Butts and Roman (2007, p. 8).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Rank</th>
<th>Mean Score: Survey 1</th>
<th>Mean Score: Survey 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Effectiveness</td>
<td>1</td>
<td>0.3</td>
<td>2.8*</td>
</tr>
<tr>
<td>AOD Assessment</td>
<td>2</td>
<td>2.6</td>
<td>4.7*</td>
</tr>
<tr>
<td>Pro-Social Activities</td>
<td>3</td>
<td>-0.02</td>
<td>1.8*</td>
</tr>
<tr>
<td>Data Sharing</td>
<td>4</td>
<td>0.3</td>
<td>2.0*</td>
</tr>
<tr>
<td>Family Involvement</td>
<td>5</td>
<td>2.1</td>
<td>3.6*</td>
</tr>
<tr>
<td>Client Information</td>
<td>6</td>
<td>2.8</td>
<td>4.5*</td>
</tr>
<tr>
<td>Systems Integration</td>
<td>7</td>
<td>1.1</td>
<td>2.8*</td>
</tr>
<tr>
<td>Targeted Treatment</td>
<td>8</td>
<td>-2.2</td>
<td>-0.7*</td>
</tr>
<tr>
<td>Resource Management</td>
<td>9</td>
<td>2.7</td>
<td>3.9*</td>
</tr>
<tr>
<td>Access to Services</td>
<td>10</td>
<td>-1.8</td>
<td>-0.7*</td>
</tr>
<tr>
<td>Cultural Integration</td>
<td>11</td>
<td>1.6</td>
<td>2.6*</td>
</tr>
<tr>
<td>Agency Collaboration</td>
<td>12</td>
<td>3.7</td>
<td>4.6*</td>
</tr>
<tr>
<td>Partner Involvement</td>
<td>13</td>
<td>5.8</td>
<td>5.7</td>
</tr>
</tbody>
</table>

Note: The 13 change indices are ordered from largest change (1) to smallest change (13). * Score changed significantly (t test) between wave 1 and wave 6 (p < .05). Amount of index change was significant (t-test) between wave 1 and wave 6 (p < .05).
Table 4
Change in survey scores across all Reclaiming Futures communities by percentage of possible improvement realized.
Butts and Roman (2007, p. 9).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Rank</th>
<th>Percent of possible improvement realized</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOD Assessment</td>
<td>1</td>
<td>28*</td>
</tr>
<tr>
<td>Treatment Effectiveness</td>
<td>2</td>
<td>26*</td>
</tr>
<tr>
<td>Client Information</td>
<td>3</td>
<td>24*</td>
</tr>
<tr>
<td>Family Involvement</td>
<td>4</td>
<td>22*</td>
</tr>
<tr>
<td>Systems Integration</td>
<td>5</td>
<td>19*</td>
</tr>
<tr>
<td>Pro-social Activities</td>
<td>6</td>
<td>18*</td>
</tr>
<tr>
<td>Data Sharing</td>
<td>7</td>
<td>18*</td>
</tr>
<tr>
<td>Resource Management</td>
<td>8</td>
<td>16*</td>
</tr>
<tr>
<td>Agency Collaboration</td>
<td>9</td>
<td>14*</td>
</tr>
<tr>
<td>Targeted Treatment</td>
<td>10</td>
<td>12*</td>
</tr>
<tr>
<td>Cultural Integration</td>
<td>11</td>
<td>11*</td>
</tr>
<tr>
<td>Access to Services</td>
<td>12</td>
<td>9*</td>
</tr>
<tr>
<td>Partner Involvement</td>
<td>13</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: “Possible improvement realized” was calculated by dividing the total difference in survey scores by the maximum possible improvement. Change indices are ordered from largest change (1) to smallest change (13), based on the percentage of possible improvement realized. For each index, a t-test was used to determine whether the amount of change was significantly different from zero.

* Amount of index change was significant (t-test) between wave 1 and wave 6 (p < .05).

Fig. 2. Change in mean index scores between six successive surveys (5 survey-to-survey changes) across 13 indices (65 total comparisons), arrayed by extent of change. Butts and Roman (2007, p. 15).

Futures had a positive effect on changes perceived by survey respondents.

Independent sample t-tests were conducted on successive survey pairs (i.e., Survey 1 compared to Survey 2, Survey 2 compared to Survey 3, etc.), and t-values were plotted from lowest to highest to determine the direction of improvement or deterioration in each index (Fig. 2). Of the 65 t-tests, 54 (83%) were positive, indicating that the mean score on a particular index was higher than the previous score. In 32 of the 54 positive instances, the difference between the first and second score was statistically significant (t > ±1.96). Eleven intersurvey changes were negative, and four were statistically significant. A Wilcoxon Matched-Pairs Signed-Ranks Test conducted on the changes across successive surveys indicated that the number of positive changes was greater than would be expected by chance alone (W = 3.97, p < .001). We note that, although the overall system change reported by respondents was positive and statistically significant, each t-test was not necessarily an independent observation, and thus, we consider these results as evidence of a strong association between Reclaiming Futures and improved participant perceptions, but we do not infer a causal effect.

The extent of change reported by survey respondents was not identical across the Reclaiming Futures communities. When assessed by the significance of the total difference from 2003 to 2006, the strongest changes appear to have occurred in Dayton (Table 5). Every one of the 13 indices of systemic change increased significantly in Dayton between the first and final survey. Other strong performers include Portland (12 significant increases), Southeastern Kentucky (10 significant increases), and the projects in Anchorage and Chicago (9 significant increases each).

6. Discussion

We find substantial evidence that the 10 communities involved in the first phase of the Reclaiming Futures initiative changed their service delivery systems in ways that were compatible with the goals of the initiative. By analyzing social networks in 8 of the 10 communities, we find that as a group these communities improved the strength of their agency networks over time. Two sites made improvements or at least stayed generally the same on all network variables, and most communities improved on a few network properties.

By analyzing stakeholder perceptions of the operations and the impact of local service delivery systems, we find that system effectiveness likely improved during the course of the Reclaiming Futures initiative. Increases were statistically significant in 12 of 13 indices of system performance. Improvements were especially dramatic in the ratings for treatment effectiveness, the use of client information in support of treatment, the use of screening and assessment tools, and overall systems integration. The findings suggest that agency processes for delivering substance abuse interventions for youthful offenders improved during the RF initiative. The survey of stakeholder perceptions was developed specifically to obtain measurements for a wide variety of factors theorized to be important indicators or outcomes that could characterize system change specifically for the RF initiative. These measures represent...
different constructs than those captured in the network measures, but the survey cannot be used directly to validate network measures, nor can network measure be used to validate the survey. However, the findings from each component can be used as complements to inform a more complete picture of the processes that occurred within and across sites as communities worked to achieve the objectives of the Reclaiming Futures model.

Of course, there are limitations to this study, many of which are endemic to a study of system change. Each local Reclaiming Futures project began the initiative with its own unique mix of resources and challenges, and its own perspectives and experiences with systemic change. The focus of their effort during Reclaiming Futures also varied considerably. Some communities worked to settle long-standing turf battles and to resolve fundamental policy conflicts that were hindering service delivery. Others addressed administrative procedures and data-sharing arrangements that affected large numbers of youth, but that were less volatile and more amenable to direct action. Each community, however, implemented a reform strategy that focused on systemic change and that sought to improve the coordination and effectiveness of the juvenile justice and substance abuse treatment systems.

Empirical analysis of system change is difficult as the unit of analysis is the system, rather than the individuals served by that system. Thus, even the largest system change initiatives like Reclaiming Futures will affect only a small number of systems, and small sample sizes limit empirical analysis. In addition, key features of the system are likely to be unobservable to researchers, and thus omitted variables that mediate effects will not be included in the analysis. As a result of these challenges, our empirical analysis had key limitations. We did not have outcome data to determine whether the Reclaiming Futures model is a successful strategy with regard to improving outcomes for drug-involved youth. In other words, we can determine using our methods whether the processes for system change were achieved as rated by agency leaders, but we cannot assess whether reported system change affected end outcomes at the individual or community level. In addition, the offender stakeholder surveys capture observers’ ratings of system change, but are likely not as accurate as direct observation of changes in service receipt. We also note that evaluating system change using network techniques is methodologically challenging. Network questions can be a burden on the respondent, and missing data can wreak havoc on analysis. Even the omission of one organization from a network can alter the composition or nature of a network (Sridharan & Gillespie, 2004).

### 7. Conclusions

Prior research has been uneven regarding how interagency partnerships implement and sustain changes in intervention systems. In part, this is because data collection about system change initiatives can be cumbersome and resource intensive, but also, systematic frameworks for measuring and evaluating system change do not currently exist. Fortunately, there is a growing body of literature focused on the study of interorganizational networks at the network level (see Provan, Fish, and Sydow, 2007, for a review) that is directly relevant to evaluating system change across systems (or within a system) that serve youth. This important area of research goes beyond an understanding of single organization change, and seeks to understand and quantify issues of whole system change: How do interorganizational collaborations evolve? How might collective outcomes be generated from structural changes to the entire network? Although much of this work has been conducted within the public management arena, the concepts and measurement techniques can be applied to evaluation of youth service systems. For instance, the work of Morrissey (1992, 1997) provides precedent for collecting network data and measuring network change within mental health systems serving youth. Social network analytic designs should be coupled with rigorous process evaluations that use performance measures and in-depth interviews of network members. The emphasis should be both on finding and measuring the appropriate factors at the baseline that will set the stage for meaningful measures of change post-baseline change, and on linking measures of system change with individual outcomes that have been hypothesized to accrue to youth.

We encourage future evaluators to work early in system change initiatives to implement performance measurement routines that can monitor how systems behave at the individual, organizational, and whole network levels. By measuring what the system did to prepare for change (inputs) and what it actually delivered (outputs), actors engaged in system reform can measure in real time how effective their efforts have been in creating the change they seek. The development and exchange of data to facilitate measurement of systems change is itself a critical component of system change. A system change initiative oriented around developing efficiencies and bridging gaps is principally about the exchange of data and other related resources. Practitioners within an organization must act in concert to achieve mutual goals, and practitioners seeking to collaborate and coordinate across systems likewise must follow the same script. Systematically assessing features of collaboration and system change through performance data, research, and practice will build the capacity to ask not only “What outcomes, both intermediate and long-term, did the partnership or initiative produce?” but also “What was the quality of the partnership itself?”

### Acknowledgements

This research was supported by a grant from the Robert Wood Johnson Foundation. Views expressed here are the authors’ alone. They do not reflect official positions or policies of the Robert Wood Johnson Foundation, John Jay College of Criminal Justice, Temple University, or the Urban Institute, its trustees and funders. The authors are grateful to Jennifer Yahner for preparing the analysis of social networks included in the study. We would also like to thank Demelza Baer, Lisa Brooks, Elissa Gitlow, Gretchen Maener, Jay Reid, Meghan Salas, Aaron Sundquist, Simon Tidd, Alyssa Whitby, and Janeen Buck Willison for their important contributions to study on which this paper is based. We also think Mark Lipsey for sharing and interpreting effect size data.

### References

