Stachowiak, S., Lynn, J., & Akey, T.. (2020). Finding the impact: Methods for assessing the contribution of collective impact to systems and population change in a multi-site study. In A. W. Price, K. K. Brown, & S. M. Wolfe (Eds.), *Evaluating Community Coalitions and Collaboratives*. *New Directions for Evaluation*, *165*, 29–44.

2

Finding the Impact: Methods for Assessing the Contribution of Collective Impact to Systems and Population Change in a Multi-Site Study

Sarah Stachowiak, Jewlya Lynn, Terri Akey

Abstract

John Kania and Mark Kramer put forward "Collective Impact" in 2011 as a framework for organizing multi-sector collaborative efforts to achieve change at scale. The collective impact theory of change posits that by establishing and implementing its five conditions, groups can achieve meaningful systems changes to create long-term gains in social and environmental conditions. While significant scale uptake has occurred, questions have remained about the degree to which collective impact, as an approach, actually works to achieve change at scale. In 2017, ORS Impact and Spark Policy Institute embarked on an evaluation effort to understand the degree to which the collective impact approach contributed to population-level change across many sites. We sought to answer this question with as much rigor as possible, without attempting to simplify the complexity of the context, the variability of implementation of collective impact, or the many interim changes needed to see the impact at scale. This chapter shares the essential methods our research team used. We do not seek to share the findings; instead, we hope that others can learn from and use these methods to continue to strengthen the sector's understanding of when, how, and why different collaborative efforts work or do not. In addition to describing the key methods, the authors will reflect on considerations, lessons learned, and

New Directions for Evaluation, no. 165, Spring 2020 © 2020 Wiley Periodicals, Inc., and the American Evaluation Association. Published online in Wiley Online Library (wileyonlinelibrary.com) • DOI: 10.1002/ev.20398

recommendations to other evaluators who might seek to answer similar questions or use similar tools and methods.

Introduction

Increased high school graduation rates. Fewer homeless individuals. Better river health. The desire for these kinds of large-scale, systemic improvements for people and places motivate many coalitions and collaborative efforts. However, there have been few rigorous evaluations assessing the contribution of large-scale collaborative efforts to these kinds of population-level changes.

In part, this is due to the mismatch between traditional approaches to assessing attribution (e.g., randomized control trials or RCTs) and the inherently complex nature of the systemic changes that collaborative efforts seek to influence, changes that are multi-faceted, non-linear, and adaptive. Traditional program evaluation approaches seek to understand the efficacy of individual programs and focus on assessing relatively discrete and fixed interventions. Large-scale collaborative efforts, in contrast, are emergent and steadily adapting, are context-dependent and responsive, and include multiple programs and activities working in mutually reinforcing ways. This complexity makes understanding how they contributed to meaningful outcomes challenging (Cabaj, 2014). Most examples and resources for effectively evaluating one such type of collaborative initiative, the collective impact approach, tend to simplify this complexity by encouraging a focus on specific initiatives within the larger body of collaborative work (Parkhurst & Preskill, 2014), an orientation necessary for ongoing learning and improvement, but one that does not help answer questions about collective impact as an overall approach for cross-sector collaboration.

In 2017, ORS Impact and Spark Policy Institute embarked on an evaluation effort to understand the degree to which the collective impact approach contributed to population-level change across many sites. We sought to answer this question with as much rigor as possible, without attempting to simplify the complexity of the context, the variability of implementation of collective impact, or the many interim changes needed to see the impact at scale. This chapter shares the key methods our research team used. We do not seek to share the findings; rather, we hope that others can learn from and use these methods to continue to strengthen the sector's understanding of when, how, and why different collaborative efforts work or do not.[1]

[1] Please review the full report (Lynn et al., 2018) for all the findings.

Collective Impact as an Approach

As coined in the Stanford Social Innovation Review, collective impact is the commitment of a group of important actors from different sectors to a common agenda for solving a specific social problem at scale (Kania & Kramer, 2011). Collective impact initiatives are distinct from other forms of collaboration in their cross-sector composition and their implementation of the five conditions of collective impact (common agenda, shared measurement, mutually reinforcing initiatives, continuous communication, and backbone support). In addition to the conditions, the approach also includes eight principles of practice that are important to achieving population-level outcomes: design and implement the initiative with a priority placed on equity; include community members in the collaborative; recruit and co-create with cross-sector partners; use data to continuously learn, adapt, and improve; cultivate leaders with unique system leadership skills; focus on program and system strategies; build a culture that fosters relationships, trust, and respect across participants; and customize for local context.

The Study

In early 2018, in light of field critiques (see Boumgarden and Branch (2013), Le (2015), Wolff (2016), and Wolff et al. (2017)) and with the support of several prominent foundations (the Annie E. Casey Foundation, the Bill & Melinda Gates Foundation, the Houston Endowment, the Robert R. McCormick Foundation, the Robert Wood Johnson Foundation, and the W. K. Kellogg Foundation), the Collective Impact Forum sought a partner to conduct a rigorous study to understand the ways in which collective impact has contributed to population change in a sample of mature collective impact efforts. Ultimately, the study examined 25 initiatives in the United States and Canada. The study defined "population change" as changes in the conditions of a target population of an initiative, which could be people or ecosystems within specific geographic areas. The research process included three phases of data collection: (1) interviews with two informants from each of the twenty-five sites and document review of available materials; (2) in-depth data collection with eight sites to conduct process tracing; and (3) in-depth data collection with three sites to explore equity approaches and outcomes (Lynn, et al., 2018).

Methods to Allow for Meaningful Cross-Site Comparisons

A vital goal of the study was to be able to make judgments about implementation and contribution to population change across an intentionally diverse set of initiatives so that we could weigh in on the efficacy of the collective impact approach overall. The twenty-five sites in the study covered a range of focus areas (e.g., river health, opioid use, teen pregnancy, juvenile justice), length of implementation (i.e., from a minimum of 3 years to more

than 20), and a breadth of geographic scopes (e.g., within a city, regional, statewide). We structured the study to establish a causal description of how the change had occurred in each site and compare across sites to identify linkages between collective impact and outcomes (Figure 2.1).

Our approach meant, of course, that the design had to have ways to standardize and make judgments about key study variables in light of the diversity. We used rubrics, qualitative coding and analysis, process tracing, structured virtual focus groups, and additional quantitative analyses focused on equity. Below we describe these methods in more detail, specifically focusing on why they were useful for a rigorous, multi-site study. Detailed descriptions of all phases and methods can be found in Appendix A of the full report (Lynn et al., 2018).

Rubrics

A rubric is a powerful tool for making judgments using a variety of types of data and across difficult to compare cases or contexts. They give structure to the analysis by providing clear criteria for judging performance or progress on specific concepts (Davidson, 2004; Oakden, 2013). Having comparable data across sites meant we could look for meaningful patterns that connected different types of changes (e.g., achieving early outcomes or systems changes) to the strength of implementation of collective impact and equity. The rubrics focused on three of the four central concepts of interest in the study: the five conditions of collective impact, population change, and equity. Their development and structure varied depending on the concept. Systems change, the fourth core concept in the study, was also explored similarly. However, the rubric did not seek to make a judgment on the strength or quality of a systems change, but rather helped to categorize each systems change that emerged in order to understand the types of changes happening across sites. The full set of indicators assessed for each rubric can be found in Appendix A in the full report (Lynn et al., 2018).

We developed *the collective impact rubric* based on guiding documents promoted by the Collective Impact Forum to establish a definition of each condition and its core characteristics (Figure 2.2). The resources used to inform the collective impact conditions rubric included Edmondson and Hecht (2014), Kramer and Kania (2011), Preskill, Parkhurst, and Juster (2013), Turner, Merchant, Kania, and Martin (2012), and Collective Impact Forum (2016). From there, we developed the indicators using previous evaluations and evaluation guides to identify what evidence would signal the presence and strength of the condition.

We used data from multiple sources to rate each indicator and condition as full/mature implementation, partial/emerging implementation, not implemented, or no data. The strength of the assessment was also rated, based on the clarity and consistency of the supporting data (Figure 2.3 provides a sample from a rubric to illustrate the process).

Figure 2.1. The causal relationships theorized across the key variables in the study.

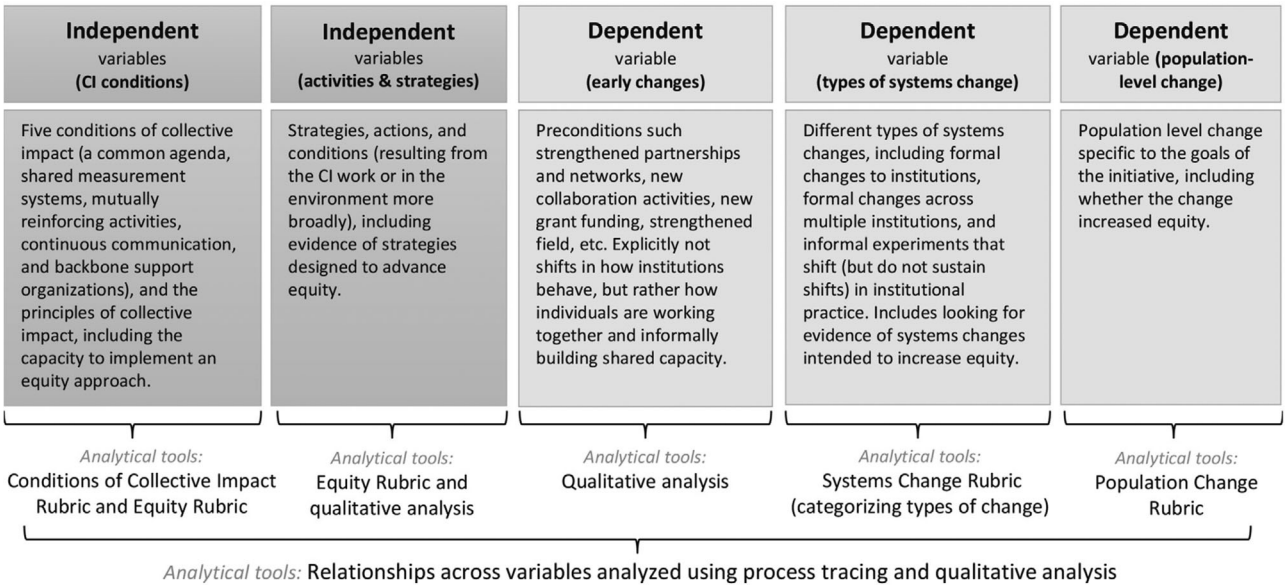


Figure 2.2. Collective impact rubric elements.

Categories	*Indicators*
BACKBONE SUPPORT	• The initiative has established a steering committee or leadership structure with responsibility and authority for governance and decision-making • The initiative has designated one or more organizations that have dedicated staff to perform backbone functions • The backbone infrastructure coordinates and supports core initiative activities such as guiding vision and strategy, convening stakeholders, supporting alignment and shared measurement practices, building public will, community engagement, and ownership, advancing policy, and mobilizing resources • Backbone staff are seen as having appropriate skills and credibility to perform backbone functions
COMMON AGENDA	• Partners and participants have a common understanding of the problem • There is an identifiable overarching goal and vision for the initiative with clearly defined boundaries and focus • Partners and participants have clearly articulated a portfolio of strategies and agreed-upon actions to that drive change • Partners are committed to a shared vision for change and have agreement upon the goals and approaches to achieving that vision
MUTUALLY REINFORCING ACTIONS	• The initiative has a collective plan of action that specifies the strategies and actions that different partners have committed to implementing • Partners are aligning their own practices and actions with initiative goals and collective action plan(s) • Working groups (or other collaborative structures) are established to coordinate activities in alignment with the plan of action • Partners hold each other accountable for implementing activities as planned
CONTINUOUS COMMUNICATION	• The initiative has structures and processes in place to inform, engage, and seek feedback from internal (collective impact partners) stakeholders, such as working groups that hold regular meetings, newsletters, or online platforms • The initiative has structures and processes in place to engage external stakeholders, such as regular meetings, websites, public convenings, public reports, and social and traditional media campaigns • Communication strategies and messages about decisions, actions, priorities, and other important aspects of the initiative are public and transparent • Communication is wide-reaching and adapted to reach a broad audience of participants
SHARED MEASUREMENT SYSTEM	• The initiative has established a set of agreed-upon indicators and data collection methods to track progress toward its outcomes • The initiative has a well-designed data infrastructure for storing and reporting data • Quality data on a set of meaningful indicators is available to partners in a timely manner • Collective impact partners use data from the shared measurement system to make decisions and establish priorities

Figure 2.3. Elizabeth River: Pull Out Box.

Common Agenda Indicator	*Rubric Rating*	*Example of Data Used to Make Judgement*	*Strength of Data Rating*
(a) There is an identifiable overarching goal and vision for the initiative with clearly defined and bounded/ actionable problem space (i.e., problem is bounded in a way that is actionable)	2 (full implementation)	Mission: "To restore the Elizabeth River to the highest practical level of environmental quality through government, business and community partnerships." The initiative defines the problem as contaminated Elizabeth River watershed due to human/ industrial activity. In initiative documentation references are made to specific problems in the river or in river areas, such as elevated levels of PCBs (polychlorinated biphenyl) in fish and crab and high levels of bacteria.	2 (clear and consistent data)
(b) Partners have a common understanding of the problem	ND (no data)	No evidence of struggle here, but no positive evidence of "common understanding" either. Interviewees believe in shared vision and buy-in among partners to goals and strategies but had less to say about holding a common understanding of the problem.	1 (vague or mixed data)

We used an iterative process to develop an *equity rubric* (Figure 2.4) that included equity in both the process and outcomes of the initiatives. Drawing first upon resources available through the Collective Impact Forum (Collective Impact Forum, 2016; Edmondson & Hecht, 2014; Williams & Marxer, 2014), and then gathering feedback from the study's partners and funders, we created indicators for the capacity of the initiative to engage equity work; how the initiative is prioritizing equity-focused actions; the degree of meaningful representation and inclusion; and the equity outcomes (system and population changes). Similar to the collective impact rubric, we reached an initiative-level assessment for each indicator and category, as well as documentation around the strength of evidence behind each rating.

The *population change rubric* included a set of descriptive codes to characterize each population change being assessed and a set of criteria for assessing its strength and validity of the population change considering the degree of reliability and validity of the change measure and data, the meaningfulness of the change, and data source.

The rubric analysis was conducted using the interviews and document reviews from the first phase. However, when Phase II data revealed additional information, the data was updated, and the final score changed for some sites. When ratings of data strength for the assessment were low, analysts followed up to collect additional data to make a judgment call.

Figure 2.4. Equity rubric elements.

Categories	*Indicators*
CAPACITY FOR EQUITY	• Backbone has necessary skills and attitudes to engage communities, develop leaders, and shift power • Initiative partners demonstrate readiness to engage in equity work through openness to dialogue and willingness to examine systemic inequity • Initiative leaders and backbone have credibility with and are trusted by local communities
EQUITY ACTIONS	• The initiative uses locally relevant and disaggregated data to identify priorities and areas for intervention • The initiative prioritizes strategies focused on addressing the disparities experienced by a focus population • Initiative solutions are shaped by community members and build on community assets and resources • Initiative partners engage in an ongoing analysis of structural inequities that drive disparities to identify systemic or root causes of inequity when developing interventions, making key decisions, or setting policies
REPRESENTATION AND INCLUSION	• The majority of leaders, implementers, and influencers are representative of the communities intended to benefit from the initiative • Resources are consistently provided to support participation of community members across multiple aspects of the initiative (e.g., meetings at convenient times/locations, bi-lingual translation of meetings, transportation/ child care, compensation for time and expertise) • Initiative makes effort to engage non-joiners and traditionally disenfranchised group

Cross-Site Qualitative Analysis

The rubrics gave us an understanding of the strength of implementation of collective impact and equity processes and the achievement of equity and population outcomes. We also collected comparable qualitative data across other variables beyond the rubrics, including coding the data into quantitative categories for types of systems changes (i.e., changes to core institutions), early changes (i.e., changes that lay the foundation for systems and policy changes, including increasing partnerships or community engagement), barriers, and facilitators.

As our goal was to generate multivariate informed themes across cases and types of variables, not summarize individual variables or tell the story of a specific site, we needed methods that facilitated pattern finding in a large volume of qualitative data across many sites. To do this, we combined

the rubric scores with the coded qualitative data using cross-tabs and Chi-Square analyses, looking for patterns in how variables significantly grouped. For example, we found a significant association between the mature implementation of a backbone and mature implementation of continuous communication (two of the conditions of collective impact). Some patterns involved just two variables; other times, multiple variables were considered together, such as looking at associations among all five conditions.

Where a pattern emerged, follow-up analysis was done to assess whether the data were robust enough to unpack the pattern and if there was a clear conceptual grounding for the pattern. If so, an analyst generated a summary of the pattern by seeking evidence confirming the pattern and evidence of any contrary or competing patterns.

Process Tracing

We used process tracing to answer the key question of the study: whether there is a relationship between the collective impact approach and demonstrable population change, or changes among people or places targeted by initiatives (e.g., rates of homelessness, river health). Process tracing is a rigorous qualitative method that explores competing hypotheses reflecting different plausible explanations of the causes of a given outcome (e.g., a population change) and rates each hypothesis' level of inferential strength (Collier, 2011; Punton & Welle, 2015).

These causal inference tests provide consistent, specific ways to assess each hypothesis based on the sufficiency and uniqueness of the data associated with the hypothesis and its rival hypotheses (Figure 2.5). Like the rubrics, process tracing gave us a tool for understanding degrees of contribution across the parts of diverse theories of change, comparatively, more robustly than through purely thematic comparison.

To implement the method, we selected eight of the twenty-five sites for in-depth engagement, based on evidence of strong implementation of at least four of the five collective impact conditions and clear evidence of population-level change. A designated research team member (the "site lead") developed a theory of how change had happened in the site based on Phase I data using the structure presented in Figure 2.1. This was reviewed and revised with the site's primary point of contact before a site visit. The site visits included: an interview with the site evaluator; two focus groups (one with stakeholders representing the backbone and governance of the initiative and a second with stakeholders representing the participants in the initiative); and a 2 hour facilitated dialogue specifically focused on reviewing and refining the theory of change, identifying alternative drivers, and assigning proportional weights to the elements of the theory that most strongly contributed to the population change.

These data resulted in a summary narrative of the contribution story, an updated theory of change noting proportional weights assigned by the

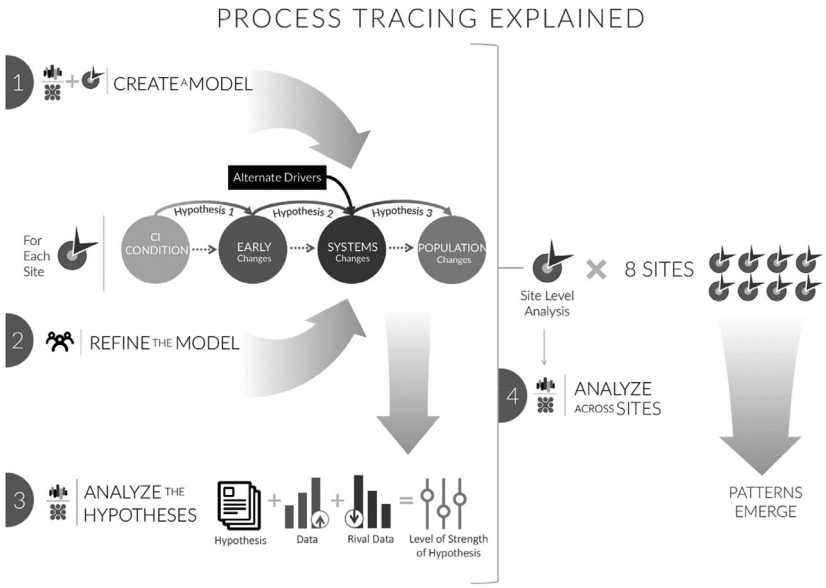
Figure 2.5. Levels of inferential strength assessed through process tracing.

Level of Inferential Strength	*Strength of Evidence*
1. The hypothesis is plausible but is neither proven or disproven.	Evidence is suggestive of a relationship, but insufficient to draw a definitive conclusion as to the contribution to the outcome relative to other rival explanations
2. The hypothesis is certain but not unique.	Evidence is sufficient to conclude a relationship exists, but not to rule out the possibility that the outcome would have also occurred due to rival explanations
3. The hypothesis is plausible and unable to be explained by a rival explanation.	Evidence is sufficient to conclude that a relationship exists and that the outcome would not have occurred due to rival explanations
4. The hypothesis is deemed to be "doubly decisive"	Evidence provides high certainty of contribution and there is no alternative explanation. This level of strength is extremely unlikely when talking about complex systems change initiatives.

stakeholders for what most strongly contributed to their sites' population change(s), and a set of site-specific hypotheses that described how the site's stakeholders understood the presumed connections between the components (e.g., which collective impact conditions led to which early changes). For each hypothesis, site leads documented available evidence and their ratings and rationale associated with the necessity and sufficiency of the hypothesis in explaining the observed change, and the plausibility of an alternative scenario in providing a stronger explanation than the site hypothesis. They then provided an overall rating of high, medium, or low for the strength of each relationship based on all the data. In addition to rating each hypothesis, site leads made a summary judgment about the strength of the relationship between the full set of hypotheses and the population-level change.

We then reviewed the site leads' data and assessed each hypothesis using the tests of inferential strength, focusing on the certainty with which you can understand the relationship as well as the uniqueness or sufficiency in fully explaining the outcome (Figure 2.7). Across sites, we analyzed the hypotheses by type (e.g., collective impact conditions to early changes), and frequencies and cross-tabs were used to identify cross-site themes and patterns and compare differences between site visit sites ($n = 8$), other

Figure 2.6. Process tracing cross-site analysis overview.



sites with population change ($n = 12$), and other sites without population change ($n = 5$) (Figure 2.6).

Equity Deep Dives

Most efforts to collectively create social change occur within systems and institutions that have systemic barriers that keep groups and individuals from benefitting equally. Just as the social sector has deepened and broadened conversations about persistent inequities and how to support a more equitable society, the role and importance of equity in collective social change efforts have also evolved. While the study examined equity across our entire study, we selected three sites to understand more deeply how equity integrates into different collective action sites. We used the Collective Impact Forum's definition: *Equity is fairness achieved through systematically assessing disparities in opportunities and outcomes caused by structures and systems and by addressing these disparities through meaningful inclusion and representation of affected communities and individuals, targeted actions, and changes in institutional structures and systems to remove barriers and increase pathways to success* (Urban Strategies Council, 2015). The sites were chosen using the following criteria: (a) evidence of an explicit, equity-focused issue area, and/or work specifically targeted to equity-focused population; (b) led by those with lived experience: either run by or significant involvement in action planning and implementation; and (c) the collaborative was

Figure 2.7. Types of hypotheses analyzed in process tracing.

Type of Hypotheses	Example Hypothesis
Collective Impact Conditions to Early Changes	Common Agenda, Mutually Reinforcing Activities (action plan and work groups) and Backbone have (1) increased trust among partners (2) created culture change among agencies (toward collaborative work), (3) facilitated development of local work; and (4) maintained high levels of political will
Early Changes to Systems Changes	The early changes (legislative champions/political will, allies, deepened relationships/trust, and public engagement) together led to the adoption and implementation of legislation and legislatively mandated implementation/oversight bodies.
Systems Changes to Population Changes	More aligned policies and practices within partner agencies/organizations have produced more targeted and higher quality services to homeless clients and increased housing placements for homeless veterans and the chronically homeless.
Overall Theory of Change Hypothesis	The extent of river clean-up and the changed relationship of the community to the water would not have occurred without the initiative.

grassroots or grasstops, not established by or composed primarily of funders, agencies, and governmental agencies.

We conducted structured virtual focus groups with diverse stakeholders from each site, including community members and leaders and frontline staff from partner and implementing organizations. In the focus groups, we asked groups to rate their initiative (using online polling software) on each of the four dimensions of the equity rubric concerning how well they felt that dimension was being implemented or achieved by their initiative. After each dimension rating, we asked the group to reflect on why they rated it as they did. They also shared thoughts on the process in which the initiative engaged to achieve the dimension, evidence of success, challenges and associated solutions, and lessons learned about infusing equity into their work.

The administration of the focus groups varied across the three sites to collect data in culturally responsive ways. For example, one grassroots initiative focused on Latino students in an urban neighborhood, a second focus group was conducted in Spanish to elicit responses from students, families, and community members who are direct intended beneficiaries of the collective. The focus group conducted with an indigenous collective in Alaska occurred as a series of story-telling segments, where the facilitators asked the group to narrate their experiences around the four dimensions

using their cultural narratives and perspectives, rather than a more objective rating and description of discrete dimensions.

Based on the data collected from Phase I and the focus groups, we completed equity rubric ratings and created descriptions of equity work in each of the deep dive sites. We then conducted thematic coding across the narratives and equity rubrics to elicit common patterns and variation across the three sites. The equity deep dive sites helped to deepen our understanding of how equity practices and outcomes occurred across sites, strengthening our ability to interpret the findings from the rubric data across all twenty-five sites. While the study design does not support conclusions about how equity capacity, action, and inclusion lead to systems and population changes, it did examine the descriptive relationships among these to tease out how they co-occur.

Lessons Learned and Recommendations for the Field

Within and across our methods, we sought to go beyond descriptive stories and examples of change in order to more deeply understand in a consistent way across sites the types of questions that are difficult to unpack in most collective impact studies due to limited data or number of sites. Below are three insights and ideas we believe researchers engaged in similar work should be aware of and explore within their work.

The Rubrics Are a Powerful Tool, Especially With Topics With a Strong Literature Basis

The cross-site collective impact rubrics were a powerful tool for describing the maturity of the collective impact implementation. The rubrics helped to avoid the problem of being overwhelmed by descriptive data across many sites and facilitated cross-site analysis by providing analytic standards for how to judge the evidence available. The data to populate the rubrics were reasonable in scope, completed from a document review, and two key informant interviews, suggesting this is a tool that could be used in a variety of studies. Although this lesson is specific to the collective impact context, similar style rubrics might work well whenever there are principles or practices that sites seek to adhere to in their implementation. Perhaps the most obvious and most straightforward next step from the study is the broader application of the collective impact rubric. It is appropriate for within site or across site studies and could also be used as a repeated measure over time. We believe the types of explicit judgments the tool makes can serve as feedback for sites that are exploring how to strengthen their collective impact implementation.

While the collective impact rubric relied upon a body of literature about the practices of collective impact that was mostly aligned around agreed-upon definitions of strong practice, the equity rubric required more

judgment calls. It is essential to understand that equity concepts may manifest differently based on culture, population, and issue area, and equity work is a complex, emergent, and iterative process where collaboratives may take steps forward in some areas, and steps back in others, depending on context and conditions at any given time. As a result, it is crucial to unpack the natural variation across sites as well as common patterns in capacity, actions, and outcomes. Additional work is needed to identify how to be responsive to unique contexts and ways in which sites act upon equity, and yet maintains comparability in the data, particularly if methods like rubrics are being utilized to strengthen the rigor of a study. Like the collective impact rubric, we believe that further testing, refinement, and development of the equity rubric on both process and outcomes could serve as a useful tool. Also, our study raised additional questions related to equity: what does it look like when collaboratives tackle the structural underpinnings of inequity versus ameliorating inequitable outcomes? Could we test the hypothesis that taking an equity approach may mean slower but more permanent change? How can the field better understand what equity approaches can look like that take into account the unique dynamics and cultures of any particular effort?

The Process Tracing Methodology Grounded the Study in Many Ways, Strengthening the Rigor and Quality of All Components of the Work

The process tracing methodology proved to be a powerful tool for establishing the causal relationships, particularly in light of the cross-site nature of the study, as well as more generally driving the study design to create rigorous, comparable cross-site data across all phases and activities in the study. The process tracing depended on in-depth and specific qualitative data, captured across multiple perspectives, to build triangulation and confidence in the findings that fed into the process tracing analysis. In other words, the work we had to do to prepare for the process tracing also ensured the data used in other parts of the study were high quality, consistent across sites, and confirmed in multiple ways. Challenges presented by the process tracing were more than offset by the benefits, but they were meaningful. In particular, the unit of analysis of the process-tracing method is the change itself, not an initiative. All the data has to help understand that change and the data needs are significant, limiting resources for other types of investigation. Additionally, the method worked well on a specific, well-defined model of collaborative process, that is, collective impact, as it allowed for comparability in implementation across the sites. Cross-site process tracing may be more challenging when sites lack this level of commonality.

Truly Understanding Equity Within a Collaborative Context Would Benefit From a More Integrated Approach

In the end, the majority of sites in our study had an emerging equity focus or no equity focus at all. And because the stronger equity sites included in the

study had not yet achieved population change at scale, sites with the most active equity practices and early equity-focused outcomes were not part of the process tracing sample. The equity deep dive with three sites was a mid-course correction to ensure we had meaningful equity learning despite this gap, but while this additional data collection was useful, in hindsight we see better ways to incorporate the equity focus: (a) more inquiry and focus on how equity was incorporated into the implementation of the five conditions; (b) building in more questions about who is being hurt, not just helped, by initiative actions and how structural inequities are considered; (c) and considering process tracing to understand achievement of equity-focused systems changes versus using the method only for documented population change. We hope to see other researchers continue to explore and advance learning around how to center equity studies of collaborative work.

Conclusion

Ultimately, we believe this study helps demonstrate the value of rigorous, cross-site analysis of complex systems change. Much was learned in the study that can be used by existing collaborative efforts to improve their work and strengthen their impact. While the investment in this type of study is significant, the value it brings site by site and particularly across sites suggests similar work is worth doing in the future. We encourage more funders to consider rigorous evaluations of their multi-site efforts to drive large-scale, collaborative change.

References

Boumgarden, P., & Branch, J. (2013). Collective impact or coordinated blindness? *Stanford Social Innovation Review*. Retrieved from https://ssir.org/articles/entry/collective_impact_or_coordinated_blindness

Cabaj, M. (2014). Evaluating Collective Impact: Five Simple Rules. Retrieved from https://www.collectiveimpactforum.org/sites/default/files/10-Evaluating%20Collective%20Impact%20-%20Five%20Simple%20Rules.pdf

Collective Impact Forum. (2016). *Collective impact principles of practice*. Retrieved from https://www.collectiveimpactforum.org/resources/collective-impact-principles-practice

Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics*, *44*(4), 823–830. https://doi.org/10.1017/S1049096511001429

Davidson, E. J. (2004). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: SAGE Publications.

Edmondson, J., & Hecht, B. (2014). Defining quality collective impact. *Stanford Social Innovation Review*, *12*(4), 6–7.

Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*, *9*(1), 36–41.

Kramer, M., & Kania, J. (2011). *Collective impact*. Retrieved from https://www.fsg.org/publications/collective-impact

Le, V. (2015). *Why communities of color are getting frustrated with collective impact*. Retrieved from https://nonprofitaf.com/2015/11/why-communities-of-color-are-getting-frustrated-with-collective-impact/

Lynn, J., Stachowiak, S., Akey, T., Gase, L. et al. (2018). *When collective impact has an impact: A cross-site study of 25 collective impact initiatives*. Retrieved from http://orsimpact.com/DirectoryAttachments/10262018_111513_477_CI_Study_Report_10-26-2018.pdf

Oakden, J. (2013). *Evaluation rubrics: How to ensure transparent and clear assessment that respects diverse lines of evidence*. Melbourne: Better Evaluation.

Parkhurst, M., & Preskill, H. (2014). Learning in action: Evaluating collective impact. *Stanford Social Innovation Review*, *12*(4), 17–19.

Preskill, H., Parkhurst, M., & Juster, J. S. (2013). *Guide to evaluating collective impact*. Retrieved from https://www.collectiveimpactforum.org/sites/default/files/Guide%20to%20Evaluating%20CI%2001.pdf

Punton, M., & Welle, K. (2015). *Applying process tracing in five steps* (Annex No. 10). Brighton: Centre for Development Impact.

Turner, S., Merchant, K., Kania, J., & Martin, E. (2012). Understanding the value of backbone organizations in collective impact. *Stanford Social Innovation Review*. Retrieved from https://ssir.org/articles/entry/understanding_the_value_of_backbone_organizations_in_collective_impact_1

Urban Strategies Council. (2015). *Equity*. Retrieved from https://urbanstrategies.org/about/equity/

Williams, J., & Marxer, S. (2014). *Bringing an equity lens to collective impact*. Retrieved from https://www.collectiveimpactforum.org/sites/default/files/EquityandCollectiveImpact_UrbanStrategiesCouncil.pdf

Wolff, T. (2016). Voices from the field: 10 places where collective impact gets it wrong. *Nonprofit Quarterly*. Retrieved from https://nonprofitquarterly.org/2016/04/28/voices-from-the-field-10-places-where-collective-impact-gets-it-wrong/

Wolff, T., Minkler, M., Wolfe, S. M., Berkowitz, B., Bowen, L., Butterfoss, F. D., & Lee, K. S. (2017). Collaborating for equity and justice: Moving beyond collective impact. *Nonprofit Quarterly*. Retrieved from https://nonprofitquarterly.org/2017/01/09/collaborating-equity-justice-moving-beyond-collective-impact/

SARAH STACHOWIAK is owner and CEO of ORS Impact, a consulting firm that specializes in evaluating the complicated, messy, and complex work that characterizes what our clients do to support meaningful, equitable social change.

JEWLYA LYNN'S independent consultancy, PolicySolve, helps social change leaders achieve their goals to make the world a better place through coaching, research, evaluation, and facilitation focused on complexity, systems thinking, and futures thinking.

TERRI AKEY is a director at ORS Impact, where she engages her passion for creating just and equitable social systems, believing that evaluative thinking and access to high-quality, relevant evidence has the power to hold system leaders accountable and help organizations change their practices to meet the needs of children, families and communities.