# Testing Effects of Community Collaboration on Rates of Low Infant Birthweight at the County Level

Adam J. Darnell · John P. Barile · Scott R. Weaver ·
Christopher R. Harper · Gabriel P. Kuperminc ·
James G. Emshoff

**Abstract** Interorganizational collaboration has become a popular strategy for addressing population health and well-being. However, evidence for its effectiveness in achieving outcomes at the population level is limited, at least in part due to a variety of methodological challenges such as reduced sample size at the population level, the availability of suitable comparison groups of communities, and study durations that are too short to detect slowly emerging outcomes. The present study addresses these challenges by retrospectively examining the effectiveness of a mature network of community collaboratives, using latent growth modeling of longitudinal change in an archival community-level outcome, low infant birthweight, and propensity score matching of comparison communities. A group of 25 Georgia counties with collaboratives targeting low infant birthweight was compared to a weighted comparison group of counties from other southeastern states, using propensity score matching. We report results of full matching methods and outcome analyses examining differences in change in county rates of low infant birthweight from 1997 to 2004 between intervention and comparison counties. Results indicated significantly smaller increases in low weight birth rates in intervention counties than in comparison counties.

Adam Darnell is currently at Casey Family Programs.

A. J. Darnell (✉)
Casey Family Programs, Atlanta, GA 30307, USA
e-mail: darnelladam@hotmail.com

J. P. Barile · S. R. Weaver · C. R. Harper · G. P. Kuperminc
Georgia State University, Atlanta, GA, USA

J. G. Emshoff
EMSTAR Research, Inc., Atlanta, GA 30307, USA

Interorganizational collaboration has become a popular strategy for addressing a wide range of problems, from teen pregnancy to cardiovascular disease. Theoretically, inter-organizational collaboration is a promising strategy due to its ability to foster relationships among individuals and organizations which had previously worked in parallel, to combine diverse perspectives on an issue and solutions, to devise comprehensive interventions targeted at multiple levels of the community, and to create strength in numbers in order to leverage resources and exert influence on external decision makers, among many other theoretical advantages (Butterfoss and Kegler 2002). In many instances, the purpose of collaboration is to create change in population-level health or behavior outcomes. However, evidence of effectiveness in doing so is lacking (Hallfors et al. 2002; Klerman et al. 2005; Kreuter et al. 2000; Roussos and Fawcett 2000), at least in part due to substantial methodological challenges inherent to this research. These challenges include the smaller size of samples of communities, the feasibility of random assignment of communities to experimental conditions, the relatively long study duration that may be required to identify slow developing changes at the community level, and the causal complexity that must be accounted for in analyses attempting to link collaboration to macro-level outcomes (Berkowitz 2001; Gabriel 2000; Roussos and Fawcett 2000; Yin and Kaftarian 1997). This study examines effects of interorganizational collaboration on a community-level indicator of child and family well-being, the rate of low weight births in Georgia counties. We illustrate a methodological approach that addresses a number of the

challenges posed by community-level research, and present results that suggest a relationship between collaboration and improvements in population rates of low weight births.

The community collaboratives in this study are part of Georgia Family Connection (FC), a statewide network of community collaboratives working to improve child and family well-being. FC collaboratives are very similar in nature to other forms of interorganizational collaboration that have been variously referred to as coalitions, partnerships, networks, etc. (e.g., Butterfoss 2007; Granner and Sharpe 2004; Roussos and Fawcett 2000; Zakocs and Edwards 2006). There is not a consensus definition of interorganizational collaboration, but the various definitions commonly identify collaboration as the interaction of diverse parties in the pursuit of a common goal (see for example: Himmelman 1996; Mattesich et al. 2001; Taylor-Powell et al. 1998; Wood and Gray 1991). In the present study, the term *community collaboration* is used to refer to a version of interorganizational collaboration distinguished by the active participation of private citizens, in contrast to membership that is determined solely by professional affiliation (Butterfoss 2007). As Bailey and Koney (1996) note, community involvement ensures that collaborative activities are germane to community needs and increases the likelihood that collaborative efforts can be sustained beyond formal organizational support.

Georgia initiated its long-term commitment to community collaboration as a strategy to improve the well-being of children and families in 1991 when 15 counties received funding to pilot the FC community collaborative approach. Over time the FC network added new counties so that by 2004 there were 157 collaboratives serving every county in the state. FC collaboratives consist of representatives from public agencies, businesses, faith-based organizations, elected officials, and local citizens. FC collaboratives serve as local decision-making bodies for their communities to improve targeted indicators of child and family well-being such as low infant birthweight, teen pregnancy, high school graduation, and unemployment. Through a community assessment each collaborative identifies the issues of relevance to their community. The FC community assessment is a strengths and needs assessment conducted by members of the collaborative, occasionally with the assistance of a hired contractor. Methods include primary data collection through focus groups, key informant interviews, and surveys, as well as secondary data sources such as Kids Count (Annie E. Casey Foundation 2009a), the US Census, and a variety of other state agency databases. Through the community assessment, collaboratives identify one or more issues, and then develop and implement strategies to address them. In this study we focus on the subset of collaboratives addressing low infant birthweight (LBW), one of the outcomes most commonly targeted by FC collaboratives.

Low infant birthweight is a strong predictor of infant mortality and is linked to a variety of subsequent health problems, such as cerebral palsy, heart and intestinal problems, asthma, and blindness as well as cognitive and social deficits such as lower IQ scores, lower educational attainment, and hyperactivity (Eichenwald and Stark 2008; Hack et al. 1995). Another reason LBW is commonly targeted by FC collaboratives is that it is largely preventable by basic prenatal care. The likelihood of women giving birth to LBW infants can be greatly reduced if mothers maintain a healthy weight, keep up-to-date on vaccinations to prevent infections, take prenatal vitamins, do not smoke, and effectively manage chronic conditions such as diabetes and hypertension (Annie E. Casey Foundation 2009b; Georgia Family Connection Partnership 2009).

FC collaboratives devise strategies that are defined as clusters of related programs, services, and activities designed and supported by a collaborative to achieve desired benchmarks for children, families and communities. As an example of an LBW strategy, several FC counties implemented an early childhood strategy that focused on supporting families with children birth to age four. This strategy included five primary services: universal contact at birth, intensive home visitation, developmental childcare, parenting education, and adult education and job training. Contact at birth and intensive home visitation included core services to improve women's health such as linkage to a medical home and health education. Referrals and linkages to prevention services during the interconception and prenatal periods were also made, including medical services, dental services, immunizations, healthy lifestyle classes and smoking cessation.

According to the FC theory of change, collaborative operations including engagement of members, governance, finance, planning, and evaluation are expected to affect low birthweight through both individual and community level pathways. Through the strategy, a collaborative may provide (or facilitate or act as a catalyst for) individual-level services (e.g., home visiting), it may facilitate changes in the system of services such that the array of services is more comprehensive and integrated (e.g., bringing home visiting providers together to reduce duplication), and it may implement activities targeted directly at the community level (e.g., by holding a legislative breakfast on LBW or running a public awareness campaign).

It is important to note that the FC theory of change does not limit the development of the strategy to any particular model of intervention. This contrasts with other models of collaboration that are more prescriptive of the method of intervention. Most prescriptive would be community trials which aim for uniform implementation of a single researcher-selected intervention model (see for example: Community Intervention Trial for Smoking Cessation: COMMIT Research Group 1995; Minnesota Heart Health

Program: Luepker et al. 1994). As a contrasting example, the Communities That Care model (CTC; Hawkins et al. 2008, 2009) strikes a balance between uniformity and flexibility by specifying that communities use a prevention science approach to identify risk and protective factors for targeted problems and choose from a menu of evidence-supported interventions available to address those factors. The model of collaboration in the present study is even less prescriptive in that FC collaboratives are not limited to the implementation of evidence-supported interventions. This naturally leads to much larger variation in the interventions implemented by FC collaboratives. We raise this distinction because it has implications for the interpretation of differences in outcomes between intervention and comparison communities.

In the present study, there is presumably such large variation within both intervention and comparison groups in terms of intervention methods (i.e., strategies) that the primary difference between intervention and comparison groups is the FC model of collaboration itself. In contrast, in a study of a model of collaboration more prescriptive in terms of intervention methods, such as the CTC model, which implements a limited set of evidence-supported interventions, intervention communities share both the model of collaboration and a more homogeneous method of intervention. In the case of CTC, studies have shown that intervention communities do in fact have higher levels of implementation of evidence-supported interventions than comparison communities (Fagan et al. 2011, 2012).

In the present study, the only consistent difference between intervention and comparison communities is the FC model of collaboration. We assume that intervention communities vary in terms of specific interventions employed to affect individual recipients of services and communities, and also that many comparison counties implemented interventions to address LBW, but the latter did so without the benefit of FC collaboratives. Thus our analyses are a test of whether FC collaboration is associated with improvements in LBW above and beyond effects of conventional services. We test a simple hypothesis, that the rate of change over time in county rates of low infant birthweight will be more favorable where there are FC collaboratives targeting that outcome, than in comparable counties in other southeastern states without FC collaboratives.

## Method

### Sample

#### Collaboratives

At the end of each fiscal year since 1997, leaders from FC collaboratives have completed a comprehensive self-assessment (different from the community assessment described earlier) that collects data on a variety of collaborative operations, including the specific outcomes targeted for the year. Targeting means that the collaborative developed a strategy to address a particular outcome. This study is focused on a subgroup of 25 counties with collaboratives that targeted LBW for a sum of at least 2 years from 1997 to 2003. This definition of the intervention offered the best combination of dosage and group size. Because targeting is measured annually, various patterns of targeting are possible over this 7-year period. The average sum of years targeting LBW during this period was 3.4 (SD = 1.5, Min = 2, Max = 7) and the average number of consecutive years targeting was 2.6 (SD = 1.6, Min = 0, Max = 7). Collaboratives may also target additional indicators. Other indicators commonly targeted by these 25 collaboratives included teen pregnancy (83 %), child abuse and neglect (73 %), and school absenteeism (71 %). Counties with collaboratives that did not identify LBW as a priority issue for 2 or more years during this time period were not included in this study so that our comparison was between collaboration targeting LBW and a 'no collaboration' comparison condition.

Just as studies of individuals report demographic characteristics of participants, we present the following characteristics of our sample of collaboratives. Based on 2003 self-assessment data, the most recent year included in the operational definition of the intervention, the average age of collaboratives was 6.8 years (SD = 2.6, Min = 2, Max = 10). The average tenure of the collaborative coordinator, the person responsible for managing day-to-day operations of the collaborative, was 3.9 years (SD = 2.7, Min = 0.1, Max = 10). The average tenure of the chair of the governing board was 1.8 years (SD = 1.4, Min = 0, Max = 6). In terms of organizational formalization, 19 of the 25 collaboratives had membership requirements (76 %), 22 had by-laws (88 %), 21 had a committee structure on the governing board (84 %), and 24 had written member roles & responsibilities (96 %). Regarding the composition of membership, 15 of 25 collaboratives included consumers of services (60 %), 12 included youth (48 %), and 22 (88 %) included business and local government, respectively.

#### Comparison Counties

In order to test the effectiveness of collaboratives in reducing LBW rates, the 25 FC counties were statistically matched with comparison counties from Arkansas ($n = 75$), Mississippi ($n = 82$), Tennessee ($n = 95$), and Virginia ($n = 135$) using propensity score methods. These states were chosen because of their geographic proximity to Georgia and the availability of county-level LBW data

**Table 1** Means for matching variables among FC and comparison counties, before and after matching

| | FC | Comparison | | Improvement | | |
|---|---|---|---|---|---|---|
| | | Before PSM | After PSM | Std. Bias* Before PSM | Std. Bias* After PSM | % improvement Std. balance |
| Total county population | 86,045 | 47,250 | 84,211 | 0.29 | 0.01 | 95 |
| % population under age 18 | 39 % | 36 % | 38 % | 0.51 | 0.03 | 94 |
| % population under 18 that is Black | 31 % | 24 % | 29 % | 0.30 | 0.08 | 72 |
| % population under 18 that is Hispanic | 5 % | 3 % | 4 % | 0.33 | 0.04 | 87 |
| % children in single-parent households | 26 % | 24 % | 25 % | 0.20 | 0.07 | 67 |
| SES composite score | 76.03 | 75.72 | 76.34 | 0.09 | −0.09 | 2 |
| Increase in total population | 15,786 | 5,954 | 16,281 | 0.38 | 0.02 | 95 |
| % increase in total children <18 years old | −2 % | −3 % | −2 % | 0.34 | −0.02 | 93 |
| % increase in Black children <18 years old | 0 % | 0 % | 0 % | 0.13 | 0.03 | 77 |
| % increase in Hispanic children <18 years old | 3 % | 2 % | 3 % | 0.30 | 0.02 | 94 |
| Propensity score | 0.12 | 0.06 | 0.12 | 0.55 | 0.03 | 95 |

\* The standardized bias estimate is calculated by dividing the difference in means of the covariate between the intervention group and the comparison group by the standard deviation (Harder et al. 2010); covariates are considered balanced if the standardized bias estimate is below 0.25 (Ho et al. 2007)

for all 8-year between 1997 and 2004 that shared a common definition of LBW (number of babies born weighing less than 5 lbs, 8 oz, divided by total number of live births).

## Measures

### Matching Variables

The propensity score model variables listed in Table 1 were drawn from census estimates (US Census 1990, 2000). All variables represent 2000 census variables with the exception of the variables reflecting changes between 1990 and 2000 values (i.e., changes in total population, the percentage of children under 18, the percentage of Black children under 18, and the percentage of Hispanic children under 18), which were created to account for population shifts over time. The composite socioeconomic status (SES) variable was drawn from a compilation of economic indicators from the 2000 census provided by the RAND Corporation (2009). Indicators of socioeconomic status included: (1) percent of adults older than 25 with less than a high school education, (2) percent of adult males unemployed, (3) percent of households with income below the poverty line, (4) percent of households receiving public assistance, (5) percent of households with children that are headed by a female, (6) and median household income. The composite SES variable was normalized on a scale from 0 to 100 on which higher scores indicate higher SES.

### Low Birthweight

Low birthweight data from 387 comparison counties and 25 FC counties was obtained from the Kids Count Data

Center (Annie E. Casey Foundation 2009a) for the years 1997 through 2004. In each year the LBW rate was defined as the total number of babies born weighing less than 2,500 g (or 5 lbs, 8 oz), divided by the number of live births in the county, multiplied by 100. The use of archival community indicators as outcomes has been recommended previously as a promising strategy for outcome evaluations of community-level interventions (Gabriel 1997).

### Analysis Procedure

Propensity scores were created by including 11 observed county characteristics in a logistic regression equation (with comparison or FC county status as the dependent variable) to assign a single summary score (the propensity score) to each of the FC and comparison counties. The resulting propensity score represents the estimated probability that each county, regardless of whether it is an FC county or a comparison county, has of being a typical FC county (values range from 0 to 1, with 1 corresponding to a county with all the characteristics of an FC county). The 11 covariates used in this analysis were chosen because they were expected to be associated with the unique social, economic and demographic characteristics of FC counties, as well as predict differences in LBW between FC and comparison counties (Hillemeier et al. 2007; Roberts 1997). The complete list of covariates accounted for in the propensity score equation is shown in Table 1.

The propensity scores were then used to match FC counties to comparison counties using a full propensity score matching procedure. Based on propensity scores, full matching creates a series of matched sets that include at least one intervention county (FC) and at least one

comparison county (Stuart 2010; Stuart and Green 2008). The cases in each set are then statistically weighted. Within each set, intervention (FC) cases are assigned a weight of 1 and comparison cases are weighted in proportion to the ratio of intervention to comparison cases in the set. Full matching, compared to classical nearest neighbor or other forms of 1:1 matching, utilizes all available comparison cases and provides an optimal balance of propensity scores between intervention and comparison cases (Rosenbaum 2002; Stuart and Green 2008; Stuart 2008, 2010; Hansen 2004). For more information on full matching procedures, see Stuart and Green (2008) and Stuart (2010). Matching was conducted using the MatchIt statistical package (Ho et al. 2007) within the R statistical software package v2.11 (R Development Core Team 2010). After matching was completed, outcome analysis consisted of latent growth modeling to determine whether change in LBW rates from 1997 to 2004 differed among FC counties compared to non-FC counties. Outcome analyses were conducted with Mplus statistical software v6.1 (Muthen and Muthen 1998–2010) using latent growth modeling techniques, with 1997 representing the intercept (coded as 0), and yearly LBW rates representing growth over time until 2004 (coded as 7).

## Results

### Matching

Table 1 presents descriptive statistics for the matching variables before and after the matching procedure. The similarity between FC and comparison counties was markedly improved after the matching procedure. Specifically, after matching, the standardized bias for six of the eleven covariates improved by more than 90 %, and all fell well below the 0.25 threshold, indicating well matched samples (Harder et al. 2010; Ho et al. 2007). The smallest improvement was observed for the SES composite variable but this mild improvement (2 %) was due to the fact that this variable was already well balanced prior to matching.

### Latent Growth Model

The primary outcome of interest was change in low weight births over the period 1997 through 2004. A latent growth model was fit to determine whether FC counties had different linear rates of change in LBW rates over the 8 year study period than the comparison counties, controlling for the numerous community contextual characteristics in the propensity score model. A full information, robust maximum likelihood estimator was employed to obtain parameter estimates and standard errors that are robust to

non-normality and missing data under the assumption that data were missing at random conditional on the covariates. Less than three percent of the cases ($n = 11$ comparison counties) were removed from the analyses due to missing data on all years, and 91 % of the remaining counties had no missing data (leaving 376 total comparison counties; 25 total FC counties). This model incorporated the weights from the full matching procedure and adjusted standard errors to account for the clustering of matched sets as well. The unconditional model for LBW rates from 1997 through 2004 specified the intercept at 1997 and included a linear factor to describe change over time. This model fit the data well [$\chi^2(31) = 33.46$, $p = 0.349$, CFI = 0.99, RMSEA = 0.01] and described a slight increase in LBW rates over time ($b = 0.14$, $p < 0.001$, CI [0.10, 0.18]). The intercept (variance = 3.67, SE = 0.62, $p < 0.001$) and slope (variance = 0.01, SE = 0.01, $p = 0.423$) were positively correlated ($r = 0.71$). To further examine the amount of variance in the slope factor, we compared this model to one in which the slope variance was fixed to zero, and found a significant reduction in model fit [$\chi^2(2) = 9.54, p = 0.008$], indicating the variation in the slope was estimable. The inclusion of a quadratic factor did not significantly improve model fit [$\chi^2(27) = 31.82$, $p = 0.24$, CFI = 0.99, RMSEA = 0.02, $\Delta\chi^2(4) = 1.28, p = 0.865$]. Accordingly, effects of FC were tested in the linear model to determine whether change in LBW rates over time in FC counties differed from comparison counties.

The conditional model included two predictors: the dummy coded intervention variable (0/1) and the SES composite score to address the possibility that this variable was not sufficiently accounted for by the propensity score. The slope and intercept parameters were regressed on both predictors. This model also fit well [$\chi^2(43) = 59.85$, $p = 0.045$, CFI = 0.98, RMSEA = 0.03]. Findings from the conditional model indicated non-significant differences between FC and comparison counties in initial LBW status in 1997 ($b = 0.63$, SE = 0.37, $p = 0.086$, 95 % CI [−0.088, 1.353]), but significant differences in the rate of change in LBW rates over time ($b = -0.13$, SE = 0.06, $p = 0.041$; 95 % CI [−0.25, −0.01]. The rate of increase in LBW rates over time was significantly smaller among FC counties than among comparison counties. The standardized effect size for the difference in mean slopes ($\delta$) was 1.48 (Raudenbush and Liu 2001), indicating that the rate of change in LBW in FC counties was approximately one and a half standard deviation units lower than in comparison counties. Applying the estimated growth rate for comparison counties to FC counties as a reflection of what would have happened without FC collaboration, these findings suggest that collaboration prevented approximately 50 low weight births across the 25 FC counties over the 8-year study period. Additionally, higher SES was
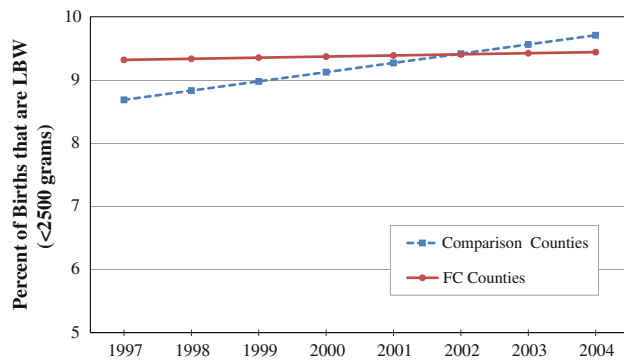
**Fig. 1** Model estimated rates of LBW among FC and comparison counties from 1997 to 2004

associated with a lower intercept ($b = -0.27$, SE = 0.02, $p < 0.001$, 95 % CI [$-0.31$, $-0.23$]) and a smaller slope ($b = 0.01$, SE = 0.01, $p = 0.004$, 95 % CI [$-0.02$, $-0.001$]). The difference between FC and comparison counties is illustrated in Fig. 1, which displays the estimated LBW trajectories for both groups.

## Discussion

Although community collaboration has become a popular strategy for addressing a range of different outcomes, reviews have found limited evidence for its effectiveness in improving population-level health and well-being (Hallfors et al. 2002; Klerman et al. 2005; Kreuter et al. 2000; Roussos and Fawcett 2000). In this study we have demonstrated a research approach that allows for examination of outcomes over a relatively long study period using a relatively large sample of communities, and we have produced evidence of a relationship between community collaboration and beneficial change in county rates of low birthweight. Linear change in county rates of low infant birthweight from 1997 to 2004 was modeled among 25 Georgia counties with Family Connection collaboratives targeting low infant birthweight, and a comparison group of 376 counties in other southeastern states statistically weighted for valid comparison. Although rates of low infant birthweight increased on average over the study period for both FC and comparison counties, the rate of increase for FC counties was significantly lower than in comparison counties, with propensity score matching and regression-based control for a variety of community contextual factors likely associated with rates of low infant birthweight and intervention group membership.

According to the Family Connection theory of change, this relationship between collaboration and county-level rates of LBW would result from changes at the individual level, via services and changes in the way services are delivered, and at the community level via changes in community awareness, policy, or funding, for example. These are the intermediate outcomes theorized to result from the FC model of collaboration which entails engaging a membership, governance, finance, planning of a strategy based on the community assessment, and evaluation to monitor progress. Due to the absence of data reflecting the nature of collaboration and services delivered in intervention and comparison counties, this study is purely theoretical in terms of the process by which an effect of collaboration would operate. We assume that comparison counties had a variety of conventional services targeting LBW, as did FC counties, so that the primary difference between groups is the value added by collaboration.

The primary threat to the internal validity of these findings is the unknown set of variables not accounted for in our propensity score model that are associated with group membership and LBW rates. In other words to the extent that we have not accounted for all of the factors associated with selection into the intervention group as opposed to comparison, the estimate of the effect of collaboration is biased in an unknown direction. Unfortunately, sensitivity analyses designed to detect this potential hidden bias (Becker and Caliendo 2007; Rosenbaum 2002) are currently unavailable using full matching and the MatchIt software program, an issue further complicated in the current investigation because our outcome is a latent slope as opposed to a point estimate such as a group mean. One of the more salient possibilities is that because all of the FC counties were located in Georgia, an unknown state-level variable not accounted for by our propensity score model could explain the difference observed between FC and comparison counties. For example, if Georgia initiated a new state policy affecting LBW during the study period, the observed difference between FC and comparison counties could be due to that policy. However we are aware of no such policy or other state initiative in Georgia, aside from Family Connection, that targeted LBW during this time period.

Another possible explanation for the observed difference in slopes between intervention and comparison counties is a difference in initial LBW levels (or correlation between change and initial status). If the amount of change in LBW is related to the initial level of LBW, and initial levels of LBW differ between intervention and comparison groups, these initial group differences could confound results. The estimated trajectories in Fig. 1 suggest that FC counties had higher initial LBW levels, but the model intercept, which represents the initial level, was not significantly related to the intervention group variable. In addition, the correlation between the intercept and slope was positive, suggesting that FC counties would tend to have larger increases in LBW were the difference in intercepts significant. Thus it is unlikely that the observed difference in slopes between FC and comparison

counties can be explained by differences in initial LBW rates.

One of the greatest challenges of research at the community-level is the causal complexity that must be accounted for in linking collaboration to community-level outcomes (Berkowitz 2001). Doing so requires clarity in both conceptual and operational definitions of collaboration and intermediate outcomes that precede improvements in health and well-being. There is considerable inconsistency in conceptual definitions of collaboration as an overarching concept (Wood and Gray 1991; Thomson et al. 2007), not to mention conceptual and operational definitions of elements of collaboration such as functioning and intermediate outputs (Granner and Sharpe 2004). In this study we have attempted to clearly characterize both the collaboratives and their theory of change. Interpretation of this analysis would be aided by more complete observation of services provided to address LBW in both intervention and comparison counties as well as the extent of and quality of collaboration occurring in both groups. Such data would illuminate the differences between the groups in terms of intervention, as well as provide possibilities for understanding mediating pathways between collaboration and observed outcomes. It should be noted, to the extent that interorganizational collaboration did occur in comparison counties we would expect less pronounced differences in outcomes. Although we did not have data to illuminate the pathways by which collaboration might makes its effect on LBW, there is a growing body of literature examining intermediate outcomes of collaboration (Brown et al. 2010, 2011; Emshoff et al. 2007; Hallfors et al. 2002; Nowell and Foster-Fishman 2011; Zakocs and Edwards 2006). Ultimately, future studies should account for both variation in the extent or quality of collaboration, intermediate outputs of collaboration such as plan quality, implementation quality, and relevant health and well-being outcomes. There is a great deal of work to be done in identification of the essential aspects of collaborative functioning and intermediate outputs across the various types of collaboration being studied.

Although measurement of the nature of collaboration and services would have been a difficult task in this relatively large sample communities, the fact that we were able to examine outcomes in such a large sample of communities is one of the major strengths of the use of archival community outcome data. This approach also allowed us to examine outcomes over a longer period of time than would typically be afforded by original data collection. The availability of the same data for comparable communities allowed for the identification of a matched comparison group.

It should be noted that because this study was conducted entirely at the community level the ecological fallacy is potentially applicable (Greenland and Robins 1994). In the case of a variable like SES, there are likely unique effects to be estimated for both the individual- and community-level variance components of SES. Because we did not model the effect of individual-level SES, our estimate of the effect of county-level SES is not to be taken as an estimate of the effect of SES on LBW at the individual level. The ecological fallacy is germane to the present study in that the cause of the observed relationship between collaboration and LBW at the county level may operate at the individual level. For instance, if the observed effects were solely due to individuals receiving services, then the observed relationship between collaboration and community-level outcomes would be spurious given that the cause existed at the individual level. This relates back to the theoretical causal chain linking collaboration to outcomes. This study is agnostic regarding the level at which a possible causal effect of collaboration is active, and we assume that effects of collaboration occur via both individual-level and community-level pathways. Thus the present findings must be interpreted as potentially resulting from both individual and community-level effects.

These limitations notwithstanding, this study provides a test of effects of interorganizational collaboration on an indicator of community well-being. Low birth weight is a strong predictor of infant mortality as well as a number of other negative health and developmental outcomes throughout the lifespan. We have produced evidence of a relationship between collaboration and desirable change in LBW. We have done so within the practical constraints of an evaluation with a limited budget. We have demonstrated how propensity score matching techniques and archival community indicators can be used to study an existing network of mature collaboratives over a long period of time in a relatively large sample of communities. We anticipate further extension of research on community collaboration to provide a deeper understanding of the mechanisms by which collaboration leads to changes in community health and well-being.

# References

Annie E. Casey Foundation. (2009a). *Kids Count Data Center* [data file]. Retrieved from http://datacenter.kidscount.org.

Annie E. Casey Foundation. (2009b). *Kids Count indicator brief: Preventing low birthweight*. Retrieved from http://www.aecf.org/~/media/Pubs/Initiatives/KIDS%20COUNT/K/KIDSCOUNTIndicatorBriefPreventingLowBirthWeig/PreventingLowBirthweight.pdf.

Bailey, D., & Koney, K. (1996). Interorganizational community-based collaboratives: A strategic response to shape the social work agenda. *Social Work, 41*, 602–612.

Becker, S. O., & Caliendo, M. (2007). Sensitivity analysis for average treatment effects. *Stata Journal, 7*(1), 71–83.

Berkowitz, B. (2001). Studying the outcomes of community-based coalitions. *American Journal of Community Psychology, 29*(2), 213–227.

Brown, L. D., Feinberg, M. E., & Greenberg, M. T. (2010). Determinants of community coalition ability to support evidence-based programs. *Prevention Science, 11*, 287–297.

Brown, E. C., Hawkins, J. D., Arthur, M. W., Briney, J. S., & Fagan, A. A. (2011). Prevention service system transformation using Communities That Care. *Journal of Community Psychology, 39*(2), 183–201.

Butterfoss, F. D. (2007). *Coalitions and partnerships in community health*. San Francisco: Jossey-Bass.

Butterfoss, F. D., & Kegler, M. C. (2002). Toward a comprehensive understanding of community coalitions: Moving from practice to theory. In R. J. DiClemente, R. A. Crosby, & M. C. Kegler (Eds.), *Emerging theories in health promotion practice and research* (2nd ed., pp. 157–193). San Francisco: Jossey-Bass.

COMMIT Research Group. (1995). Community intervention trial for smoking cessation (COMMIT): II. Changes in adult cigarette smoking prevalence. *American Journal of Public Health, 85*(2), 193–200.

Eichenwald, E. C., & Stark, A. R. (2008). Management and outcomes of very low birth weight. *New England Journal of Medicine, 358*, 1700–1711.

Emshoff, J. G., Darnell, A. J., Darnell, D. A., Erickson, S. W., Schneider, S., & Hudgins, R. (2007). Systems change as an outcome and a process in the work of community collaboratives for health. *American Journal of Community Psychology, 39*(3–4), 255–267.

Fagan, A. A., Arthur, M. W., Hanson, K., Briney, J. S., & Hawkins, J. D. (2011). Effects of Communities That Care on the adoption and implementation fidelity of evidence-based prevention programs in communities: Results from a randomized controlled trial. *Prevention Science, 12*, 223–234.

Fagan, A. A., Hanson, K., Briney, J. S., & Hawkins, J. D. (2012). Sustaining the utilization and high quality implementation of tested and effective prevention programs using the Communities That Care prevention system. *American Journal of Community Psychology, 49*, 365–377.

Gabriel, R. M. (1997). Community indicators of substance abuse: Empowering coalition planning and evaluation. *Evaluation and Program Planning, 20*(3), 335–343.

Gabriel, R. M. (2000). Methodological challenges in evaluating community partnerships & coalitions: Still crazy after all these years. *Journal of Community Psychology, 28*(3), 339–352.

Georgia Family Connection Partnership. (2009). *Improving infant health: Addressing low birthweight in Georgia*. Retrieved from Georgia Family Connection Partnership website: http://fcn.gafcp.org/lbw/LBWcompendium.pdf.

Granner, M. L., & Sharpe, P. A. (2004). Evaluating community coalition characteristics and functioning: A summary of measurement tools.". *Health Education Research, 19*(5), 514–532.

Greenland, S., & Robins, J. (1994). Ecologic studies-biases, misconceptions, and counterexamples. *American Journal of Epidemiology, 139*, 747–760.

Hack, M., Klein, N. K., & Taylor, H. G. (1995). Long-term developmental outcomes of low birth weight infants. *Future of Children, 5*(1), 176–196.

Hallfors, D., Cho, H., Livert, D., & Kadushin, C. (2002). Fighting back against substance abuse: Are community coalitions winning? *American Journal of Preventive Medicine, 23*(4), 237–245.

Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association, 99*(467), 609–618.

Harder, V. S., Stuart, E. A., & Anthony, J. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods, 15*(3), 234–249.

Hawkins, J. D., Brown, E. C., Oesterle, S., Arthur, M. W., Abbott, R. D., & Catalano, R. F. (2008). Early effects of Communities That Care on targeted risks and initiation of delinquent behavior and substance use. *Journal of Adolescent Health, 43*, 15–22.

Hawkins, J. D., Oesterle, S., Brown, E. C., Arthur, M. W., Abbott, R. D., Fagan, A. A., et al. (2009). Results of a type 2 translational research trial to prevent adolescent drug use and delinquency: A test of Communities That Care. *Archives of Pediatrics and Adolescent Medicine, 163*(9), 789–798.

Hillemeier, M. M., Weisman, C. S., Chase, G. A., & Dyer, A. M. (2007). Individual and community predictors of preterm birth and low birthweight along the rural urban continuum in central Pennsylvania. *The Journal of Rural Health, 23*(1), 42–48.

Himmelman, A. (1996). On the theory and practice of transformational collaboration: Collaboration as a bridge from social service to social justice. In C. Huxham (Ed.), *Creating collaborative advantage*. London: Sage.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199.

Klerman, L. V., Santelli, J. S., & Klein, J. D. (2005). Editors' commentary: So what have we learned? The Editors' comments on the coalition approach to teen pregnancy. *Journal of Adolescent Health, 37*(Suppl 3), S115–S118.

Kreuter, M. W., Lezin, N. A., & Young, L. A. (2000). Evaluating community-based collaborative mechanisms: Implications for practitioners. *Health Promotion Practice, 1*(1), 49–63.

Luepker, R. V., Murray, D. M., Jacobs, D. R., Jr, Mittelmark, M. B., Bracht, N., Carlaw, R., et al. (1994). Community education for cardiovascular disease prevention: Risk factor changes in the Minnesota Heart Health Program. *American Journal of Public Health, 84*(9), 1383–1393.

Mattesich, P., Murray-Close, M., & Monsey, B. (2001). *Collaboration: What makes it work. A review of research literature on factors influencing successful collaboration* (2nd ed.). Saint Paul, MN: Amerst H. Wilder Foundation.

Muthen, B. O., & Muthen, L. K. (2010). *Mplus statistical analysis with latent variables (Version 6.1)*. Los Angeles: Muthen & Muthen.

Nowell, B., & Foster-Fishman, P. (2011). Examining multi-sector community collaboratives as vehicles for building organizational capacity. *American Journal of Community Psychology, 48*, 193–207.

R Development Core Team. (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Version 2.11. Vienna, Austria.

RAND Corporation. (2009). Center for population health and health disparities data core [data file]. Retrieved from http://www.rand.org/health/centers/pophealth/data.html.

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*, 387–401.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birthweight in Chicago. *American Journal of Public Health, 87*(4), 597–603.

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.

Roussos, S. T., & Fawcett, S. B. (2000). A review of collaborative partnerships as a strategy for improving community health. *Annual Review of Public Health, 21*(1), 369–402.

Stuart, E. A. (2008). Developing practical recommendations for the use of propensity scores: Discussion of 'A critical appraisal of propensity score matching in the medical literature between 1996 and 2003' by Peter Austin, Statistics in Medicine. *Statistics in Medicine, 27*(12), 2062–2065. doi:10.1002/sim.3207. discussion 2066–2069.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21. doi:10.1214/09-STS313.

Stuart, E. A., & Green, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology, 44*(2), 395.

Taylor-Powell, E., Rossing, B., & Geran, J. (1998). *Evaluating collaboratives: Reaching the potential*. Madison, WI: University of Wisconsin-Extension.

Thomson, A. M., Perry, J. L., & Miller, T. K. (2007). Conceptualizing and measuring collaboration. *Journal of Public Administration Research and Theory, 19*, 23–56.

US Census Bureau. (1990). Census of population and housing [data file]. Washington DC: US Census Bureau. Retrieved form: http://factfinder.census.gov.

US Census Bureau. (2000). Census of population and housing [data file]. Washington DC: US Census Bureau. Retrieved from: http://factfinder.census.gov.

Wood, D. J., & Gray, B. (1991). Toward a comprehensive theory of collaboration. *Journal of Applied Behavioral Science, 27*(2), 139–162.

Yin, R. K., & Kaftarian, S. J. (1997). Introduction: Challenges of community-based program outcome evaluations. *Evaluation and Program Planning, 20*(3), 293–297.

Zakocs, R. C., & Edwards, E. M. (2006). What explains community coalition effectiveness? A review of the literature. *American Journal of Preventive Medicine, 30*(4), 351–361.