How does the form and structure of collaborative management relate to environmental outcomes?

Presented at: APPAM 2013 Fall Research Conference, Washington, D.C.

Tyler Scott[1]
Ph.D. Candidate
University of Washington

Abstract

In this paper, I seek to address the question of *whether –and how– collaborative governance improves environmental outcomes*. I use a representative watershed quality data series, the EPA's National Rivers and Streams Assessment (NRSA) and Wadeable Streams Assessment (WSA), in conjunction with a watershed management regime database produced and coded for this analysis, to analyze the relationship between collaborative governance and watershed quality for 357 watersheds. To test this relationship, I employ a hierarchical linear regression modeling (HLM) that links specific collaborative policy features (not solely the presence of a collaborative group, but rather what management role(s) the group serves, group outputs, group membership, and other relevant characteristics) relate to water quality and watershed health. The overarching research question for this project is *how does the form and structure of collaborative management relate to environmental outcomes?* **While I find that collaborative groups are strongly associated with environmental improvements in a watershed, it is not necessarily clear what accounts for this predicted impact.** I find limited evidence that groups engaged in policy making and management are more effective than groups that serve as coordinative bodies or information forums, and that having a dedicated coordinator generally increases a group's predicted impact on environmental conditions. However, I also find mixed evidence concerning stakeholder diversity, technical advisory groups, group codification, and goal formalization. The central takeaway is that policy scholars and practitioners need to think more deeply about why we believe that collaborative groups are an effective vehicle for service delivery and how such delivery can be improved.

[1] tscott1@uw.edu; 253.632.3362

Introduction

"Collaboration" and "collaborative governance" are normatively popular concepts that have been widely employed in environmental policy applications worldwide. Collaboration has been shown to enhance knowledge acquisition and foster belief change amongst stakeholders (e.g, Leach et al. 2013) and generate funds and support for alternative policy measures when problems are too diffuse or difficult to address via regulation (e.g., Margerum 2011). However, most of what we know about the effects of collaborative governance comes from process evaluations (assessing the quality of the collaborative process itself) or intermediate outcome evaluations (assessing the achievement of non-tangible outcomes such as knowledge acquisition or tangible outcomes such as project funds) (Carr et al. 2012). We still know very little about the effect of collaboration on resource management outcomes (Carr et al. 2012; Koontz and Thomas 2006). As a result, there is also a deficit of rigorously tested theory that would enable policy makers to wield collaborative governance as a strategic, context-appropriate policy tool to achieve resource management goals.

In this paper, I use one of the most common applications of collaborative governance,the management of watersheds and river basins (Sabatier et al. 2005; Margerum 2011), to explore my research questions. Using a representative watershed quality data series, the EPA's National Rivers and Streams Assessment (NRSA) and Wadeable Streams Assessment (WSA), in conjunction with a watershed management regime database that I have produced and coded, I analyze the relationship between collaborative governance and watershed quality for 450 watersheds. I argue that if collaborative governance is to be environmentally beneficial, then we need to view it as specific, context-appropriate policy tools that are more –or less- warranted in particular applications. In other words, we need to design, and implement collaborative governance just we design and implement individual tradable fishing quotas, regulations governing water pollutant release, and conservation easements. Thus, the overarching research question motivating this analysis is: *to what extent does the form and structure of collaborative management relate to environmental outcomes?*

In what follows, I first describe how this analysis fits within—and builds upon—the extant literature. Specifically, I embed my analysis within the policy tools framework developed by Salamon (2002). Next, I outline my hypotheses, taking care to describe the theoretical rationale on which they are based. I then detail my data collection process and coding scheme, as well as the data I use to model environmental outcomes. These data, from both primary and secondary sources, provide rich covariates that allow me to tease out the relationship between collaborative watershed management and environmental outcomes. This is followed by a description of the model I employ to test my hypotheses. In describing the model, I also explain why hierarchical linear modeling is appropriate given my data and research questions. I then present and discuss the results of my analysis. Finally, I conclude with a discussion of broader implications and future directions for this research. In the next section, I address the theoretical rationale for this work:

Rationale

As I briefly outline above, I believe that collaborative management is appropriately viewed as a "policy tool," i.e., "an instrument or means used to address public problems" (Salamon 2002, 2). I find that the current theory and evidence about the use of collaborative management in environmental policy is insufficient in this regard. In terms of providing empirical guidance for design and implementation, we still know very little about the effect of collaboration on resource management outcomes (Carr et al. 2012; Koontz and Thomas 2006). While the increasing diversity of policy tools (including grants, tradable permits, payments for services, and public-private partnerships) can provide flexibility and opportunity, it also presents a challenge, as policy makers must decide not only whether to act, but which tool to select. Each policy tool possesses "its own decision rules, rhythms, agents, and challenges" (Salamon 2002, 6), and thus must be carefully chosen, designed, and implemented. For instance, founding legislation and group charters accord specific management responsibilities, set membership and representational standards, and frame decision-making criteria (e.g., majority vs. consensus) for collaborative management groups. Currently, the literature cannot speak definitively to the tradeoffs –in terms of function and effectiveness—inherent to these design choices.

As a wide and growing body of literature attests, many scholars are interested in studying the role and impact of collaborative governance. Collaboration has been shown to enhance knowledge acquisition and foster belief change and trust amongst stakeholders (e.g, Leach et al. 2013; Lubell et al. 2010), and to raise funds and support for alternative policy measures when policy problems are too diffuse or difficult to address via regulation (e.g., Margerum 2011). However, most of what we know about the effects of collaborative governance comes from case studies and survey methodologies that evaluate the process (assessing the quality of the collaborative process itself) or intermediate outcomes (assessing the achievement of non-tangible outcomes such as knowledge acquisition or belief convergence) of collaborative management (Carr et al. 2012). Findings about the environmental outcomes caused by collaborative management, however, have been more limited. Several case studies have identified environmental gains resultant from the use of collaborative groups in management, and also collaborative groups that have not achieved meaningful environmental change (e.g., Carr et al.

2012; Margerum 2011). These studies do provide causal narratives about the environmental effects of the collaborative groups studied. However, from a policy perspective, it is difficult to operationalize these results, since they do not necessarily produce operable findings that can directly inform whether –and how– policy makers use collaborative groups as a policy tool in other settings. Thus, there is a deficit of rigorously tested theory that would enable policy makers to wield collaborative governance as a strategic, context-appropriate policy tool in order to achieve resource management goals such as improved or maintained water quality.

The empirical premise of my research is that publically supported collaborative management efforts should be structured and implemented on the basis of demonstrated environmental impacts. Currently, such efforts are hindered by a lack of data connecting collaborative management to environmental impacts. In order to test this connection in a policy-relevant fashion, it is first important to conceptualize the choices policy makers face when choosing to support or engage in collaborative management. I propose a typology that reflects key managerial decisions, such as whether a collaborative group engages in information sharing, planning, or policy implementation, and the level of inclusiveness or diversity of group membership (e.g., inter-agency workgroup or multi-sector group that involves private and/or non-profit actors). The current collaborative governance literature contains numerous typologies, but these largely concern the policy scale at which a group operates instead of functional and structural characteristics of collaborative management itself.

Perhaps the most prominent and useful existing typology is that of Margerum (2008; 2011), who distinguishes between the institutional scales on which collaboration occurs; however, applying collaborative management as a purposeful policy tool requires what might be termed a within-level approach. In other words, while Margerum's typology speaks to the institutional context(s) in which collaborative occurs, it does not distinguish between the choices policy makers face when electing to implement collaborative governance within a given level. Similarly, Moore and Koontz (2003) characterize groups in terms of composition (e.g., agency-based or stakeholder based), but I am purposely interested in groups that are inherently agency-based to on extent or another. Other scholars (notably Ansell and Gash [2008] and Emerson et al. [2012]) have developed theoretical frameworks for collaborative governance that speak to issues of institutional design. However, these frameworks do not distinguish between specific policy choices but rather identify key variables, such as participatory inclusiveness and stakeholder incentives, which mediate outcomes. Thus, a theoretical perspective of collaborative management as a policy tool requires a new typology that accounts for functional characteristics such as group responsibilities and structural characteristics such as group membership in a way that reflects the concrete choices policy makers face.

Hypotheses

My primary hypothesis is that *the characteristics of collaborate watershed management groups impact the effect a group has on water quality outcomes*. Based on this broad hypothesis I test four dimensions of collaborative watershed management: (H1) the level of management responsibility accorded to the collective ('Group Responsibility'); (H2) the inclusivity of representation in the group ('Stakeholder Representation'); (H3) group formalization ('Formalization'); and (H4) the primary source(s) of group funding ('Funding Source'). These variables are empirically relevant because conflicting evidence obscures how participation incentives and institutional design impact the effectiveness of collaborative management (Ansell and Gash 2008; Heikkila and Gerlak 2005; Newig and Fritsch 2009). In the reminder of this section, I provide a brief overview of each sub-hypothesis and orient each within the collaborative governance literature. Before proceeding, however, it is important to note that these hypotheses are not mutually exclusive; any combination or subset of these characteristics can conceivably be shown–or not shown—to enhance the impact of a collaborative group on water quality.

Group Responsibility (H1) contrasts groups that serve as coordinating bodies or engage in outreach or monitoring from groups that participate in planning and on-the-the ground projects, and further from groups that have actual management responsibilities such as rulemaking, enforcement, and policy implementation. Ansell and Gash hypothesize that incentives to manipulate and act co-optively are checked in situations in which actors expect to engage in ongoing cooperation (2008, 560). This is in keeping with the broader literature on collection action, which finds that a longer time horizon is necessary to foster norms of reciprocity (e.g., Ostrom 2000). An empirical example of this within collaborative watershed management is the responsibility accorded to a given watershed group. Some groups are formed for relatively short-term tasks, which, even when repeated, do not require ongoing cooperation. For instance, a group that simply meets once a year to discuss what member organizations are engaging in does not require the level of cooperative behavior necessary for a group that shares ongoing management responsibility. Generally, the literature holds that increasing the intensity of interactions (e.g., from information sharing to planning to joint implementation) requires greater stakeholder engagement and investment (Margerum 2011; Sabatier et al. 2005b; Wondolleck and Yaffee 2000): such heightened activity should result in improved environmental conditions.

Stakeholder Representation (H2) considers collaborative governance bodies on a spectrum in terms of whether a group is comprised solely of local governments (cities, counties, and special districts) versus including higher-level institutions (e.g., state and Federal agencies) and external organizations such as tribes, businesses, agricultural interests, non-governmental organizations (NGOs), universities, and technical advisory groups. There is conflicting evidence about the efficacy of increasing stakeholder representation and gaining greater input from a wider array of experts: On one hand, some find that incorporating non-state actors leads to more ecologically rationale decision-making (Dryzek 1997; Smith 2004), improved compliance (Sabatier et al. 2005), and more effective implementation (Burby 2003; Carlson 1999). However, ttempting to incorporate the interests and knowledge of all relevant stakeholders, however, can also result in diluted plans and policies that reflect the lowest common denominator of consensus (Coglianese 1999; Koontz et al. 2004).

Formalization (H3) reflects the extent to which a group possesses structural characteristics such as a dedicated coordinator or group bylaws, as well as the extent of mission formalization (mission statement, itemized goals, or itemized objectives). Groups achieve greater environmental gains with increased resource support generally (Curtis and Byron 2002; Parker et al. 2010; Yaffee et al. 1996), but it is less clear how specific expenditures, such as hiring a dedicated coordinator or developing a group constitution, alter the environmental impact of a group. Groups that have a dedicated coordinator are hypothesized to have greater impact, as the coordinator can increase administrative support and ease group tensions (e.g., Imperial 2005; Huxham and Vangen 2000; Margerum 2002; Susskind and Cruikshank 1987; Susskind et al. 1999). Similarly, bylaws are hypothesized to enhance group stability and regulate function. Lastly, increased plan and objective formalization is hypothesized enhance group effectiveness (in terms of improved water quality) by enabling groups to better assess their efficacy and better focus their efforts (Innes and Booher 1999; Margerum 2011; Wondolleck and Yaffee 2000)

Finally, Funding Source (H4) refers to primary sources of group funding, namely local, state, or Federal. The literature speaks to the necessity of "achieving 'buy in'" (Ansell and Gash 2008, 560) amongst participants even when collaborative governance is mandated. This raises a series of interesting questions for policy makers about how one might foster such an attitude amongst stakeholders. One empirical difference that inductively emerges from this research is that of group funding sources. Many watershed groups are funded by grants and other support stemming from state agencies and Federal sources such as Clean Water Act funds or the Natural Resources Conservation Service. In numerous other cases, watershed collaborative groups are more greatly supported by local municipalities and districts, area stakeholders, and private sector organizations. I hypothesize that groups with greater local support will be more strongly associated with improved water quality because participants in such groups have more "skin in the game." In other words, participants commit their own resources to the group, and thus have a vested interest in group function.

Model

The use of a hierarchical multilevel model (see Gelman and Hill 2006; Raudenbush and Bryk 2001) is crucial to answering my research question for several key reasons. In this section, I first describe the theoretical rationale for using a multilevel approach. I then describe how a multilevel model is particularly appropriate given the analysis context at hand. Following this discussion, I detail my empirical approach and modeling assumptions. Finally, I present model specifications.

Given the connectivity of watersheds and ecosystems, a sample site (on a specific reach of a specific stream) is obviously not wholly independent of nearby sites (whether sampled or not); thus, one would expect sites from the same stream or watershed to be more similar to each other than two geographically disparate sites. Likewise, water condition samples taken in multiple time periods from the same site are likely to be correlated as well. A standard regression approach for addressing this lack of independence amongst samples would be to fit an indicator (i.e., a fixed effect) for each site or for each basin. However, there are only two periods available for each of the 357 repeated samples contained in the WSA and NRSA data, and in many cases there are only one or two sample sites within a given river basin. This means that the conditional group mean for each site or basin is likely to be poorly estimated given the small sample size.

In contrast, the multilevel model takes into account the estimation uncertainty associated with each group-level adjustment (Gelman 2006). Instead of fitting a group adjustment based solely on the conditional within-group mean (as does a traditional fixed-effects model, which effectively serves to treat each group as its own distinct dataset), a multilevel model shrinks the group-level adjustment towards the sample mean as the size of the group decreases. In other words, as the within-group sample size decreases, the model places more credence upon the whole sample estimate, and vice versa. This avoids overstating differences between groups, since the multilevel model takes account of the higher potential for the within-group means in such cases to be driven by non-representative outliers. Accordingly, for data in which individual observations are nested within higher-level groupings, a multilevel model produces more reasonable inferences than does a classic fixed effects model (Gelman 2006).

A hierarchical multilevel model is also necessary for this analysis because it facilitates simultaneous analysis of variance at the sample level, sub-basin level, and state level. Fixed effects do not allow for further inference on a grouping variable, since this variation is "unmodeled" (Gelman and Hill 2006). A multilevel approach allows me to: (1) examine the effect of sample-level, sub-basin level, and state-level covariates; and (2) test for cross-level interactions, such as the relationship between particular collaborative management characteristics and state governance. Fixed effects models cannot achieve this type of analysis, because a group-level variable becomes collinear with the grouping indicators. Essentially, a hierarchical multilevel model accounts for expected correlation between repeated measures and proximate sites while still preserving the ability to model higher level (state and sub-basin) effects.

Modeling watersheds raises an additional analytical challenge in that watersheds are nested within both physical and administrative boundaries, yet these grouping factors are themselves not nested. Watersheds are defined according hydrological relationships; thus, they sometimes overlap with state or county boundaries, but typically do not. Any two sample pairs could be in the same state, but different watersheds, or in the same watershed, but in different states. Accordingly, I fit a cross-classified multilevel statistical model (Gelman and Hill 2006; Raudenbush and Bryk 2001) that groups observations both by 47 states (excluding Hawaii, Alaska, and Florida[2]) and by 148 4-digit Hydrologic Unit Code (HUC4) sub-basins.[3] Figure 1 provides a visual demonstration of how individual sample sites are cross-classified within two

[2] Most streams in Florida are coastal and thus tidally influenced, thus excluding them from the criteria for the WSA. Accordingly, Floridian streams were less likely to be selected for an original WSA sample, and even less likely to be selected for a follow-up sample under the NRSA. As a result, the 357 repeat observations do not include Florida.

[3] The Hydrological Unit Code (HUC) taxonomic system has numerous levels, from 2-digit to 12-digit identification numbers denoting hydrological regions of increasingly smaller scale. Thus, there are numerous HUC levels within which a given sample site is nested; HUC4 and HUC8 designations are among the most commonly employed analytical scales. I elect to group samples within HUC4 "sub-regions." An HUC4 sub-region encompasses several HUC8 "accounting units," which are the primary unit in which watershed data are reported. For instance, Puget Sound in Washington state is one HUC6 (1711), as is the Lower Yellowstone River (1010) and the Potomac River (0207). Grouping by HUC4 accounts for likely correlation between streams in a given basin, while allowing for each group to have an assortment of streams that are managed collaboratively and those that are not. Since most HUC8 accounting units have just one sample site (sampled under both the WSA and NRSA), there is no effective way to have treatment and control groups at that smaller level. Further, having several more observations per group facilitates more accurate estimation of inter-group differences.

different second level groupings (sub-basin and state). Variance is then modeled at each of the three levels shown in Figure 1.

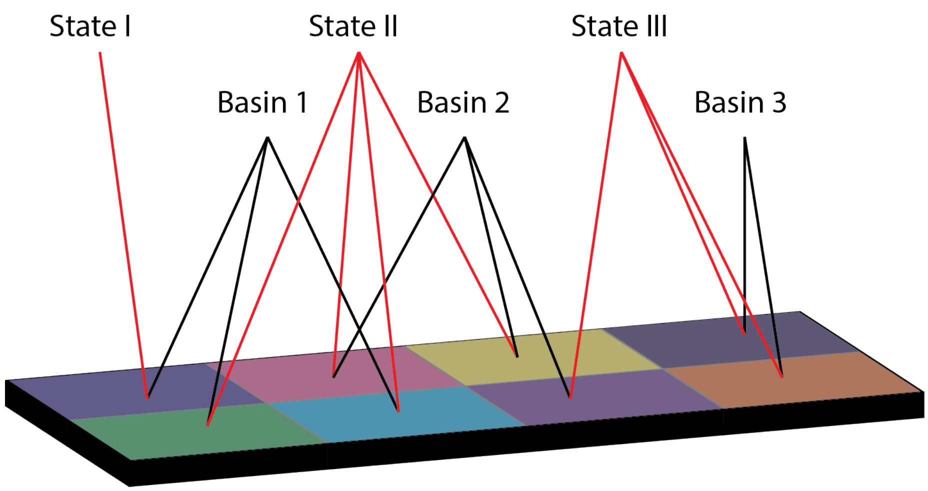

Figure 1: Graphical Presentation of Cross-Classified Model

My approach for modeling the association between collaborative management and water quality is to treat collaborative management as a policy intervention, much as we might treat a scholarship offer to a student or a new training program at a welfare-to-work site. I identify 105 unique groups across the 357 sample sites included in both the WSA and NRSA. Since many groups encompass more than one sample, however, another way of summarizing this is that I observe 259 samples (out of 714 total samples) for which there is an active collaborative management group (defined as existing prior to the sample year). The model effectively contrasts this treatment group (streams managed under a collaborative regime) with a comparison group of streams not associated with any sort of collaborative management structure. Along with estimating the direct effect of an active collaborative management group on water quality, I estimate how various group features and characteristics affect the predicted impact of a group using a series of interaction terms. I am able to examine how the predicted effect of a collaborative group varies according to the types of responsibilities given to a group or according to the types of stakeholders included. These interaction terms provides a more meaningful –and empirically grounded—interpretation, since the effect of any specific management characteristic should rightfully be expressed through the presence of the collaborative management regime and not independently.

Further, there is a temporal component to this analysis as well. The various outputs of a collaborative group such as plans or joint projects would not likely have an immediate effect; instead, it is likely that such an effect would take time to be realized. Thus, I employ several different specifications for the presence of an active group: (1) a binary indicator of whether or not a collaborative group is active in the watershed; (2) a binary indicator of whether or not a collaborative group has been active at least five years prior to the sample year; (3) a binary indicator of whether or not a collaborative group has been active at least ten years prior to the sample year; and (4) a continuous variable of the number of years that a group has been active in the watershed prior to the sample year.

At the first level of the model I estimate a water quality index score for individual stream-year *i* in sub-basin *w* and state *s*:

$$Y_i = \alpha_{w[i]} + \lambda_{s[i]} + \sum_k \beta_k Collab_{ki} C_i + \sum_l \delta_l Stream_i + \tau_i + \epsilon_{iws}$$

where Y_i represents the dependent variable, a given quality metric (e.g., nitrogen level) for sample *i*. Accordingly, $\alpha_{w[i]}$ represents the conditional intercept estimate for *i* given that it is watershed *w*; similarly, $\lambda_{s[i]}$ represents the conditional intercept estimate for *i* given that it is in state *s*. The summation term that follows represents a vector of estimated effects associated with collaborative watershed management characteristics 1 to *k* associated with observation *i* ($Collab_{ki}$), conditional on the presence of an active group C_i. Next, δ_l represents a vector of control parameters for observation-level stream characteristics 1 to *l* ($Stream_{lji}$). Finally, τ_i represents the effect of sample year for sample *i,* and ε_{iws} represents the random error associated with observation *i* in basin *w* and state *s*.

At the sample-year level outlines above, I fit controls for non-agricultural and agricultural disturbance, sample year, and mean wetted width of the stream. The disturbance metrics are intended to take account for factors that might result in a significant difference between the WSA and NRSA samples, such as the development of a stream-side property, fencing livestock from the waterway, or other proximate changes. Mean wetted width and sample year are intended to account for climatic factors such as rainfall that might cause water quality changes between samples.

The conditional sub-basin-level impact, α_w, is the dependent variable for the second model level:

$$\alpha_w = \alpha_0 + \sum_m \pi_m Watershed_{mw} + \mu_w$$

in which α_0 represents the average outcome for across basins and π_m represents the estimated effect of watershed characteristic *m*, including physical characteristics such as size and ecoregion, watershed stressor characteristics such as percent of watershed that is urbanized, population density, and road density, and management characteristics such as EPA region, state monitoring and enforcement metrics, and whether the watershed crosses a state or national boundary. Basin-level random error is denoted as μ_w. At this level, I fit watershed-level control variables for total watershed area, site elevation, percent of land area that is wetlands, percent of land area that is urban, that is wetlands, and that is forested, population density, and road density (mean road length per watershed square mile).[4] Each of these variables is log-transformed to account for a left-skewed distribution and achieve a more normal distribution. These data were calculated for the WSA, but not the NRSA, thus necessitating fitting these covariates at the watershed level. Given that no two samples are more than eight years apart, one would not expect a great deal of change in any case. Thus, the aforementioned variables are intended to act as controls for the overall context in which a sample was taken, while variables such as the human disturbance index (describe previously) account more directly for potential differences in water condition. Additionally, I fit controls for structural characteristics such as Strahler stream order and whether the watershed crosses a state or national boundary.

The conditional state-level impact, λ_s is the dependent variable for the other non-nested second level of the model:

$$\lambda_s = \lambda_0 + \sum_n \gamma_n State_{ns} + \eta_s C_{s[i]} + \mu_s$$

in which γ_n represents the estimated effect of state characteristic *n*, such as state monitoring and enforcement performance, η_s represents the modeled effect of an active collaborative group in state *s*, and μ_s represents state-level random error. For this alternative second (state) level, I will control for yearly spending by each state environmental agency and the number of inspections, violations, and enforcement activity by year and state. However, due to the continuing Federal government shutdown, these data are not available and will be included at a later date.

To evaluate the utility of the multilevel model and verify the extent to which variation

---

[4] In each case there are about 10 observations that are unrecorded; for each of these missing data points, I calculate the mean value for the given variable within the same ecoregion and impute this mean value for watersheds in the same region.

does occur at the second model levels, I compute the intraclass correlation coefficient (ICC). Since I fit two second-level grouping factors, state and sub-basin, the ICC is calculated by dividing the summed group level variance for both state and sub-basin by the sum of the group and individual level variances from the base model with no predictors:

$$\frac{\sigma_\alpha^2 + \sigma_\lambda^2}{\sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_y^2}$$

If all members within each group, the ICC equals 1 (since there is perfect correlation within each group); in such case, grouping obviously would add a great deal of explanatory power to the model. Conversely, if observations within a group are not correlated at all (thus, grouping adds no explanatory power to the model), the ICC equals 0 (Gelman and Hill 2006).

For the six dependent variables fit in this analysis, Table 1 presents the ICC for each base model, in which the dependent variable is modeled solely as a function of the sub-basin grouping indicator:

Table 1: Intraclass Correlation Coefficients

	Phosphorus Level	**Nitrogen Level**	**Turbidity Level**	**Benthic MMI**	**Riparian Cover**	**Fish Cover**
ICC	0.446	0.630	0.395	0.371	0.510	0.283

These ICC scores indicate that grouping by state and sub-basin lends a great deal of explanatory power to the model. For instance, almost 45% of all observed variation in total phosphorus level and almost 49% of all observed variation in riparian cover can be explained by the state and sub-basin grouping indicators. Even for in-stream fish cover, which has the lowest ICC value, state and sub-basin can explain more than a quarter of all total variance. This speaks to the necessity of employing a multilevel model for this analysis.

In discussing hierarchical multilevel models, it is important to note that the standard heuristics applied to fitting parameters in ordinary least squares regression and similar (e.g., logistic) models, statistical significance, is inappropriate for determining which indicators to leave in and which to leave out (Gelman and Hill 2006, 271). For instance, the model(s) includes a grouping indicator for each sub-basin, not just the indicators found to be statistically significant. The reason is that I am interested not in examining the difference between groups, but rather in generating the best possible estimate, even if that comes at the cost of precision for many (non-significant) groups. What is of interest, however, is whether a given variable is a meaningful source of variation. For instance, 77 of the 105 unique groups identified have a dedicated coordinator, while only 28 of these groups have identified bylaws. In terms of objective formalization, 61 have a mission statement, 31 have itemized goals, and 13 have specific, measurable objectives. Out of eight potential stakeholder categories, 23 groups have three or fewer such members, 24 have exactly four types, 28 have five types, and 30 have six or more. Around half (55) of these groups draw a significant portion of funds from local organizations, 47 draw a significant portion of funds from state governments, and 28 receive a significant level of Federal support. One caveat is that specific "cells" representing particular combinations of variables are rather sparse. Table 2 presents an example of the 105 groups divided in terms of objective formalization, presence of a dedicated coordinator, and receipt of local funding:

Table 2: Objective Formalization x Local Funding x Dedicated Coordinator

	Local Funding < 30% of Total		**Local Funding > 30% of Total**	
	No Coordinator	Coordinator	No coordinator	Coordinator
Mission Statement	3	8	2	18
Itemized Goals	6	26	14	15
Specific Objectives	1	6	2	4

As Table 1 demonstrates, for example, there are very few groups that have no coordinator and also receive a small proportion of their funds from local organizations. While the multilevel model mitigates many small sample size issues, I nonetheless fit covariates related to each hypothesis separately so to avoid overfitting. Further, in my analysis I emphasize broad level distinctions rather than comparisons between particular combinations of factors.

Dependent Variables

The data I use to assess water quality outcomes come from two national surveys, the Wadeable Streams Assessment (WSA) and the National Rivers and Streams Assessment (NRSA). The WSA, conducted in 2004-2005, sampled 1392 stream sites that were randomly selected from all streams of a given size within a ecological region. In other words, the sampling was stratified by ecological region and stream size.[5]

This probability-based design was thus intended to allow for generalization by ecoregion and EPA region about ecological condition, using stratification to generate a representative sample and using weighting to account for length and other relevant factors. The key point here for my purposes is stream selection was conducted independently of management variables. In other words, the presence and/or type of collaborative management regime that governs a stream played no role in selection. This presents a unique opportunity for empirical research since most

[5] The WSA surveyed only perennial, wadeable streams. Perennial refers to streams that flow year round under conditions of normal precipitation. The WSA sampling protocol is stratified by Strahler stream order. “Wadeable streams,” i.e., those that can be sampled without using a boat, are generally considered to be of orders 1 through 5. However, Strahler ordering does not directly correspond to stream size; rather, the Strahler protocol orders models streams as directed graphs, analogous to a tree. Ordering proceeds in reverse from bottom to top, thus a “leaf” stream, i.e., one that has no tributaries, is of Order 1. The Ohio River is an 8th order stream, the Mississippi River is a 10th order stream, and the Amazon River in South America is a 12th order stream. The sample was also stratified by the 9 (of 15 total) Omernik North American Level I ecoregions that occur in the continental United States, such as the Great Plains and Mediterranean California.

research on collaborative governance selects on either the independent (management characteristics) or dependent (outcome) variables.

The NRSA, conducted in 2008 and 2009, employs a similar design, but expands the sample frame to non-wadeable rivers.[6] In total, 1924 sites were sampled at random (within strata by size and ecoregion). As with the WSA, the NRSA's probability-based design means that every stream or river had a known probability of being selected, which eliminates the potential for coverage error or non-response bias. The NRSA study design incorporates a revisit of sites sampled for the WSA, with 357 original WSA sites being re-sampled in 2008-2009. These 357 sites are what I use in this analysis to conduct a longitudinal analysis. Each sample was conducted using a line-and-transect procedure that standardized data collection across all sample sites.

The WSA and NRSA assess the ecological condition of each site according to a series of measurements of chemical stressors, metrics of physical condition, and biological indicators. These different variables provide a holistic view of stream condition. From these data, I select six variables to model as representing water quality and stream health: total phosphorus content and total nitrogen content (chemical stressors caused by human activities such as mining or agriculture), water turbidity and in-stream fish habitat (physical indicators reflect more proximate habitat destruction), and indices of riparian vegetation and benthic community abundance (biological indicators of condition).

Total phosphorus and total nitrogen content are both measured in absolute terms, using micrograms per liter (ug/L) as units. Turbidity is measured in Nephelometric Turbidity Units (NTUs), using a tool called a nephelometer, which gauges the amount of light reflected by the

[6] Generally of Strahler stream orders between 6 and 10.

particles in water. In the WSA and NRSA data, most sites fall between In-stream fish cover and riparian cover are both calculated using line-transect surveys which calculated the summed areal proportion of each cover type. For instance, to calculate fish cover the surveyor assesses coverage at specific points in a 10 meter by 20 meter littoral plot. These data are then used to estimate the areal proportion of the reach that contains natural cover for fish. Because this metric is a summation of the proportion of the reach that is covered by several different kinds of cover, including boulders, large woody debris, and overhanging vegetation, this value can be greater than 1. In the data used for this analysis, sites range in value for the variable from 0 to 2.58. Similarly, because riparian cover is a summation of the proportion of the streamside riparian area that is covered by canopy, mid-layer, and ground-level vegetative cover, this value can be greater than 1 as well. Sites range in value for this metric from 0 to 2.18 in the data.

The benthic condition index is more complicated. To assess benthic condition, the WSA and NRSA generate an index for macroinvertebrate assemblage by assessing "least disturbed" sites in each ecoregion, using these sites as a baseline to estimate the "expected taxonomic composition of an [macroinvertebrate] assemblage in the absence of human stressors" (EPA 2013; Hawkins et al. 2000) (also conditional on natural gradients such as elevation or stream size). Of course, there are numerous ways to assess the condition of a macroinvertebrate community, including abundance, composition, diversity, and various submetrics related to particular taxa. Further, the appropriateness and significance of these various metrics can differ by region. Thus, for each of the 9 ecoregions within which sampling was stratified, a particular subset of 6 benthic community metrics were chosen upon which to generate a macroinvertebrate multimetric index (MMI) for each ecoregion (each individual metric is scored on a 1 to 10 scale, after which all six metrics are summed and then normalized to a 0 to 100 scale).[7]

Independent Variables

In order to develop a watershed management database, I examine (1) legislative documents that allocate management responsibilities and funds to groups; (2) group reports, mission statements, membership lists, and constitutional documents; and (3) watershed management plans (specifically the portion of each plan that discusses the use and role of public involvement) for each of the 357 watersheds that were sampled for both the NRSA and WSA. In very few cases are the majority of these sources available for a given watershed, so a primary challenge is to apply a uniform coding scheme to diverse sources.

I proceeded to systematically coding the data as follows. The coding process for each watershed began at the EPA's "Surf Your Watershed" site for the HUC8 designation associated with the observation.[8] The page provided background information including the state(s) and county(ies) with land area in the watershed, the primary watershed name, and well as links to various monitoring websites and in some cases local watershed organizations. Armed with this background knowledge, I then proceeded to search for the documentation described above. All sources used to develop these data are available from the author and published in an appendix. I also employed geographic location data present in the WSA and NRSA datasets to examine the specific location of a sample site; since some collaborative groups pertain to specific streams or sub-basins, this was necessary to posit whether a given observation could appropriately be scored

[7] For instance, in the Xeric Ecoregion (comprised of the Great Basin, much of Southern California, and the Intermountain West), the MMI incorporates metrics for Non-insect % Distinct Taxa, % Individuals in Top 5 Taxa, Scraper Richness, Clinger % Distinct Taxa, EPT Richness Distinct Taxa, and Tolerant % Distinct Individuals. In total, there are 21 different metrics that are part of the MMI for at least one ecoregion.
[8] http://cfpub.epa.gov/surf/locate/index.cfm

as under the influence of a given collaborative governance body. This multi-source approach, taking advantage of the various resources available on state and Federal agency websites and databases, is quite similar (though expanded) to the approach used by Moore and Koontz (2003) to identify and survey watershed groups in the state of Ohio.

In determining whether a watershed is considered for the purpose of this analysis to be managed collaboratively, I include only groups in which at least one governmental entity participates. Since the focus of this research is on the use of collaborative governance as a public policy tool, I am only interested in groups to which a public-sector entity devotes time and resources (since the ultimate question is whether such time and resources are being used effectively or might best be devoted elsewhere). This coding strategy proves inclusive, with only two prevalent types of watershed-oriented organizations being left out: (1) county resource conservation districts; and (2) local citizen groups that engage in watershed advocacy. Again, these types of organizations are not of interest in this particular study because I specifically focus on instances in which a public entity has chosen to devote resources towards collaborative governance.

Variables of interest are coded as follows:

Dedicated Coordinator: Groups were coded '1' if the group does have a designated coordinator and "0" otherwise. This variable does not reflect the coordinator's FTE nor the presence of group chairs, presidents, or other officers which are ubiquitous across watershed groups but do not carry administrative responsibilities.

Objective Formalization: Group objectives are assigned one of three possible values: (1)

“mission statement”: a broadly conceived sentence (or paragraph) that provides a general statement about the impetus and aims of the group; (2) “goals”: itemized, but unspecific, tenets that the group proclaims to be aiming for (e.g., “I. Improve water quality in river; II. Increase awareness about environmental behavior in community.”); and (3) “objectives”: itemized statements that outline specific actions intended by the group and/or specific metrics by which the group is able to measure its output or outcomes (e.g., “Fund local restoration projects” or “Reduce the level of stream bank erosion”). These data are ordinal, in the sense that a group’s specific objectives implicitly represent goals, and goals implicitly represent a mission statement; in other words, this variable represents an increasing level of specificity and formality of a group’s mission. This variable is modeled as series of factors in the regression model, with mission statement as the reference category. Primary sources for these data include group bylaw documents, charters, and group websites.

Inclusiveness: As specified above, the baseline requirement for a watershed being coded as having a “collaborative management group” is that the group includes a public institution as a member. Thus, the “null value” for a group’s inclusiveness is a group that is comprised solely of local governmental representatives (as specified by published membership rosters, group bylaws, and annual reports). Groups are scored for the presence of tribes, businesses, local stakeholders (e.g., advocacy organizations), non-governmental organizations (e.g., Nature Conservancy), universities or colleges, agricultural interests, Federal agencies, and state agencies. A group receives either a “1” (present) or “0” (absent) reflecting membership by each other type of organization. For instance, any group that incorporates a tribe or tribal organization receives a “1” for “Tribal Presence,” and any group that incorporates a local business or representative of a local business association receives a “1” for “Business Presence.” These values are then summed. Thus, if a group is constituted solely from representatives of local government, tribes, and the business community, then said group’s score for the number of stakeholder types included is a 2.

Funding Source: Group annual reports and founding documentation are used to code for primary funding sources. For each of three categories (Federal, state, and ‘local’), a group for which at least 30% of their funds come from a given source are assigned a ‘1’ for that variable (thus, a group can have one, two, or three coded funding sources). Many collaborative groups are funded via a combination of local funds and grants from state and Federal agencies. Thus, I employ 30% as the threshold because I believe it is important to have a coding scheme that allows a group be recorded as having significant funding from two or even all three sources. Consider a hypothetical group, with a $1 million total budget, which receives $400,000 from the EPA, $350,000 from the state through rotating CWA grant funds, $300,000 from local organizations, and $50,000 from private donations. Simply coding this group as being Federally funded would greatly belie the true distribution of funding. One might expect that this group is duly influenced by its local, state, and Federal funding. I seek to capture this potential influence in the model.

Group Bylaws: Groups for which a foundational document codifying responsibilities, membership requirements, group procedures, and similar constitutional details is published are given a ‘1’ for the presence of group bylaws. All other groups receive a ‘0’ for this variable.

Responsibility Level: In order to develop a comprehensive coding scheme for the types of responsibility policy makers accord to a collaborative group, I inductively identify seven general categories of tasks that emerge from the data: Planning, management, outreach, monitoring, coordination, projects, and education. Transaction costs theory (see Coase 1960) speaks to the monetary and non-monetary costs associated with inter-firm (or inter-organization) exchange. Cooperative efforts such as those described above require a combination of empirical resources, such as time, transportation, and contracting costs, and intrinsic resources, such as trust and social capital (Putnam 2000). In the framework of transaction costs theory, collaborative group activities such as joint policy implementation are more intensive than activities such as information sharing because they entail greater transaction costs (Margerum 2007; Wondolleck and Yaffee 2000).

Consider, for instance, the difference between simply attending a quarterly meeting in which organizations discuss their activities and having to coordinate daily operations with other organizations. The formal and informal costs associated with an informational meeting are clearly much less than those associated with joint operations. Thus, of the group activities I identify, I code coordination and outreach to be the least intensive level of group responsibility. These codes represent groups that exist as information sharing forums, hold periodic informational events, and allow for informal consultation amongst members. While I anticipated that coordinative groups would receive a designation of limited responsibility, outreach activities emerge as a common group activity in my data collection and coding. I observe many collaborative groups that put on various forms of outreach events, such as a public awareness both at a local festival or holding a watershed awareness day; in such cases, collaborative activity is limited to the scope of these activities, and requires no ongoing collaboration or internal changes within participating organizations. Planning, programs (such as restoration activities or ongoing anti-pollution programs), and ongoing monitoring conducted by groups are coded as medium intensity. These activities entail greater cooperation amongst members and greater resource expenditure by participants, but do not entail the same level of effort –or risk—as does the highest responsibility level I code, for groups that engage in joint management. These activities include permitting, rulemaking, enforcement, and policy implementation. Obviously, groups can and do engage in activities of varying responsibility level. However, I presume that if policy makers support a collaborative group to which they give management responsibilities, license to engage in lower-intensity efforts as well is implicit. Thus, each group is coded according to the highest intensity level at which it operates.

The nature of the data I use in this study are likely to raise two primary concerns, which I address here. First, questions may arise about the process of identification. My analysis finds that groups vary considerably in terms of their "presence" in grey literature (e.g., agency reports) and on the internet. Some group website contain an archival section from which I am able to access documents such as yearly reports and older documents, or a specific page which references staff or organizational members. For other groups, I am forced to use a more deductive approach. For instance, a group report might contain a reference to the group coordinator, or a group resolution might be co-signed at the bottom by group members. These data would then be used to code for the presence of a dedicated coordinator and to record membership of different stakeholder types. This heterogeneity increases the potential for Type II error in which I conclude that a group does not exist (or more likely) overlook a specific group characteristic simply because a given document or textual reference is not found or is not available. However, if: (a) Type II error is present; and (b) collaborative groups do improve environmental quality, then a failure to identify group presence or group characteristics will serve to shrink the estimated difference between the null and alternative categories. Thus, the resultant estimates are likely a lower bound on the difference between categories.[9]

Second, my data coding approach is similar to that of qualitative document analysis (QDA) (Altheide et al. 2008), often used in political science. Since QDA involves the qualitative coding of textual sources for meaning, precision and impartiality are primary methodological concerns (Guba and Lincoln 1985). Precision refers to the replicability of the analysis. In many forms of content analysis, precision is thus assessed using inter-coder testing. The intensive, voluminous nature of this data collection process prevents me (at this time) from using a multi-coder approach. Perhaps the most significant challenge is that the task of coding in large part consists of *searching* for a relevant piece of data, rather than *interpreting* a set of observations that are common across all watersheds. Similarly, my single coder approach also raises the concern of impartiality or objectivity. Both objectivity and precision are constant concerns in research based upon interpretation. However, by carefully describing and presenting the process by which I draw conclusions, I attempt to provide an "audit trail" (Platt 2006) that allows the reader to vette my approach. Thus, APPENDIX I presents the coding protocol I apply to each textual resource. This provides an overview of the analytical process I apply to each data source. Likewise, I adhere to the recommendation of Guba and Lincoln (1994) to provide full access to data so that my findings can be replicated and verified. While the nature of the database makes inclusion within this manuscript unfeasible, I make available the data sources that I employ (including group websites, plans, reports, etc.) associated with each assessed watershed. These are available by email using the contact information above. Next, I present the results of my

[9] Alternatively, this analysis could be viewed as an "easy case" for collaborative watershed management if groups that have a stronger documented presence are also those groups that have the most impact (not out of the question, given that such groups are likely more munificent and active on average). If this is the case, then a failure of this analysis to identify significant environmental impacts associated with collaborative management would not change if groups were wholly identified.

analysis.

Results

First, I present the results of the restricted model, in which no collaborative group variables are modeled except for whether or not a publically supported collaborative group is active in the watershed (Table 3).[10] The models in Table 3 use no time-lag for an active group; thus, any watershed for which a group is formed at any point prior to the sample year is considered to have an "active group." For instance, a one-unit increase in road density is predicted to increase the total phosphorus level by 19% (exp[0.170] = 1.19, since phosphorus level is log-transformed). This is a significant increase, but it is important to note that the median value for road density in the sample is 1.09, so a 1-unit increase is a very large increase that one would anticipate should have a large corresponding effect.

Table 3: Baseline Covariates, Model With No Time-Lag For Active Group

	Phosphorus Level[11]^ (-)	**Nitrogen Level^ (-)**	**Turbidity Level^ (-)**	**Benthic MMI (+)**	**Riparian Cover^^ (+)**	**Fish Cover^^ (+)**
Area[12]^	0.119*	0.103**	0.139*	1.432*	-0.016$^{\cdot}$	-0.010
Elevation	-0.309***	-0.146*	-0.254**	1.166	-0.044**	0.019
% Urban^	0.044*	0.034*	0.001	-0.354	-0.006	0.000
% Wetlands^	0.051**	0.062***	0.031	-0.378	-0.002	0.004
% Forest^	-0.103***	-0.138***	-0.066$^{\cdot}$	1.714***	0.036***	0.009$^{\cdot}$

[10] The six dependent variables used to model the effect of collaborative management are largely incommensurate in terms of scale and magnitude. One commonly used approach is to standardize each variable so that effects can be compared (since effects are estimated in z-scores). However, model goodness-of-fit is a constant priority. An analysis of standardized residuals, Q-Q plots, and DFFITS values for various model transformations reveals that the optimal approach for nitrogen level, phosphorus level, and turbidity level is to log-transformed each variable (since each is negatively skewed), while for riparian cover and in-stream fish cover, the best fitting model is achieved by taking the square root of the dependent variable (benthic community health is best fit without transformation). Transforming these variables better approximates a normal distribution and improves model fit. While using different transformations is less than ideal, as it complicates the interpretation and comparison of estimated effects, my primary interest in this analysis is examining whether collaborative groups are associated with statistically significant environmental improvements. Thus, I prioritize model fit over interpretability.

[11] Note that the dependent variables are not uniform in directionality. For phosphorus level, nitrogen level, and turbidity, a decrease represents an environmental improvement, whereas for benthic community abundance, riparian cover, and fish cover, an increase represents an environmental improvement. In order to reduce confusion and preserve conceptual understanding of these variables, I elect to keep them 'as-is' and simply note which direction represents and environmental improvement with either a (-) or a (+) sign.

[12] In 100's of square miles

Road Density$^{\wedge}$	0.033	-0.018	0.035	0.081	-0.006	-0.012*
Population Density$^{\wedge}$	-0.007	0.098***	0.045	-1.393***	-0.002	-0.007
Agricult. Disturb.$^{\wedge}$	0.081***	0.051***	0.062**	-0.535*	-0.015***	-0.008**
Non-Ag Disturb.$^{\wedge}$	0.019	0.030*	-0.016	-0.032	-0.003	0.008**

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Generally, the results of these baseline models are similar to what might be expected. Road density, for instance, is positively related to total nitrogen level, total phosphorus level, and stream turbidity, and negatively related to benthic condition (the relationship to fish cover and riparian cover is statistically insignificant). Agricultural disturbance is strongly positively related to increased levels of nitrogen, phosphorus, and suspended solids (turbidity), and strongly negatively related to the riparian cover, in-stream fish cover, and benthic condition. Colloquially, one might interpret *increased* levels of nitrogen, phosphorus, and suspended solids and *decreased* vegetation, fish habitat, and benthic abundance as "bad for the environment"; thus, these results are commensurate with what one might expect from agricultural activity proximate to streams.

As an example of how the group-level effects are estimated, Table 4 shows the intercept adjustments fitted for a single HUC4 (1708, the Lower Columbia River), and state (Montana) (since each model fits a random effect estimate for state, ecoregion, and HUC4, a full presentation of each estimate for each stream quality metric is prohibitive):

Table 4: Examples of modeled HUC4 and state effects

	HUC4: Lower Columbia River	**State: Montana**
Phosphorus Level$^{\wedge}$ (-)	-0.664	-0.178
Nitrogen Level$^{\wedge}$ (-)	-0.013	0.665
Turbidity Level$^{\wedge}$ (-)	-0.621	-0.381
Benthic MMI (+)	-0.015	0.064
Riparian Cover$^{\wedge\wedge}$ (+)	-0.032	-0.228
Fish Cover$^{\wedge\wedge}$ (+)	0.000	0.413

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted

These examples demonstrate that streams in Montana generally have less turbidity and phosphorus, and less riparian cover than the general sample population. Likewise, streams in the Lower Columbia River Basin have a lower level of chemical pollutants than the general sample population. Interestingly, the model for in-stream fish cover failed to find any substantive difference amongst HUC4s in terms of in-stream fish cover; thus, the intercept adjustment for each HUC4 is near zero. Each baseline model also allows the effect of an active group to vary by state; in other words, "Active Group" is modeled rather than simply fit as an unmodeled (i.e., fixed) effect. This provides an overview of how the impact of a collaborative watershed management group is predicted to vary by state (Table 5):

Table 5: Modeled Effect of 'Active Group' on Sample of States

	Phosphorus Level^ (-)	**Nitrogen Level^ (-)**	**Turbidity Level^ (-)**	**Benthic MMI (+)**	**Riparian Cover^^ (+)**	**Fish Cover^^ (+)**
AL	-0.37	0.02	-0.45	-0.13	-0.18	0.06
CA	0.61	-0.15	0.53	0.17	-0.35	-0.07
CO	0.03	-0.03	0.26	0.19	0.29	0.04
KS	-0.14	0.05	-0.51	-0.03	-0.23	-0.02
KY	-0.01	0.07	0.12	0.04	-0.19	0.15
MN	0.25	0.04	0.60	-0.06	0.33	-0.03
PA	-0.25	0.10	0.56	-0.14	-0.22	-0.20
UT	0.74	-0.01	0.30	-0.40	-0.29	-0.39
WI	-0.27	0.02	-0.24	0.13	-0.36	0.25

^ variable log-transformed, ^^ dependent variable square rooted

Table 5, which provides a sample of the modeled effect of 'Active Group' by state on each stream quality metric, has a relatively simple interpretation: a stream with an active group in Utah, for instance, is predicted to have an increased level of phosphorus and turbidity, a decreased level of riparian cover, and an increased level of fish cover and benthic community abundance. The primary utility of modeling this effect as differentiated by state is not to facilitate comparison of the efficacy of various state programs (since watershed groups just as much within states as they do between states), but rather to control for the fact that groups might be created and supported for different reasons in different states. In particular, some states might be more likely to form groups to restore poor quality streams, whereas others might form groups to protect currently threatened streams. Thus, allowing the total effect of an active group to vary by state adjust for the context in which groups operate.

Next, I proceed to test for the effect of an active collaborative watershed group. Table 6 shows the estimated effect of an active collaborative watershed group by several different models. Each model has the same specifications and contains the same covariates as those described above, but each model employs a different time lag for determining the threshold at which a group is modeled as being "active."

Table 6: Model With Baseline Covariates, No Time-Lag For Active Group

Active Group	**Phosphorus Level^ (-)**	**Nitrogen Level^ (-)**	**Turbidity Level^ (-)**	**Benthic MMI (+)**	**Riparian Cover^^ (+)**	**Fish Cover^^ (+)**
0 Year Lag	-0.227	-0.256*	-0.266`	-2.531	0.005	0.045*
5 Year Lag	-0.274`	-0.275**	-0.405*	-1.022	0.007	0.052`
10 Year Lag	-0.276	-0.154	-0.129	-1.412	0.027	0.064`
Total Duration	-0.026`	-0.019*	-0.032*	-0.085	0.002	0.006**

^ variable log-transformed, ^^ dependent variable square rooted
` $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results of Table 6 speak generally to a strong relationship between the presence of a collaborative watershed group and reduced nitrogen and phosphorus levels, as well as increased in-stream fish cover. For instance, groups that have been active at least five years are predicted to reduce total phosphorus in a stream by 24% (exp[-0.274] = 0.76) and turbidity by 34% (exp[-0.405] = 0.67). The results for fish cover are positive and statistically significant across almost all models, while the results for nitrogen content, phosphorus content, and turbidity are significant except across only groups that have been active for at least 10 years. The effect on riparian cover is consistent and positive, indicating that groups are associated with improved riparian condition, but in no case is the parameter statistically significant. The estimates for the model using duration of group presence is quite different in magnitude because the marginal effect for these terms is per an additional year of group presence, rather than whether a group is present at all. For instance, the multilevel model predicts that each additional year a collaborative group is active in a watershed reduces stream turbidity by 3% (exp[-0.032] = 0.97) and decreases nitrogen level by 2% (exp[-0.019] = 1.02). Finally, it is interesting to note that the effects generally become larger in magnitude as the threshold for an active group is increased (particularly from zero years to five years); this indicates that groups do take time to have an impact. While the estimates jump around some as the threshold is increased, this is likely do to the fact that increasing the threshold reduces the population of groups analyzed, thus producing a more highly variable estimate simply due to reduced sample size. The most notable change is with respect to benthic condition. In the discussion section below, I elaborate on why the four metrics for which significant effects are found might be more strongly associated with collaborative group presence than riparian cover and benthic abundance.

Building off of these baseline models, I next proceed to test how various characteristics of collaborative watershed management affect the predicted effect of collaborative group presence. It is important to emphasize that what I am modeling then is how these different variables of interest, such as the presence of a dedicated coordinator or the level of management responsibility accorded to a group, alter the predicted impact of an active collaborative watershed management group. The estimates associated with the basic, binary 'Active Group' parameter in Table 4 represent a conditional mean estimate for the presence of a group; what I am interested in, however, are how specific group characteristics make a group more or less successful at improving various aspects of stream condition. These effects can be measured by using a series of interaction terms that model the variance associated with the ‘Active Group’ parameter as a function of the management characteristics of interest.

H1: Inclusiveness and diverse representation in collaborative watershed management results in better water quality outcomes.

Table 7 shows the estimated effect of incorporating other jurisdictions in watershed management group, again fitted using several different specifications of an “active group.” The variable ‘TRANS-BOUNDARY GROUP’ refers to whether or not an active watershed group has representation from the jurisdiction opposite of the sample site (for instance, the British Columbia Ministry of the Environment). While a lack of congruity between administrative and environment boundaries is often held up as a driver of environmental mismanagement (e.g., Karkkainen 2002), Table 7 fails to speak to any environmental improvement associated with involving other states or provinces in a collaborative watershed management group. No effect, from any model specification, is found to be statistically significant; further, the direction and magnitudes of these estimates do not appear to evidence even a statistically insignificant positive environmental effect. If transboundary representation does improve water quality, one would expect that the coefficient to be negative for phosphorus level, nitrogen level, and turbidity, and positive for benthic condition, riparian cover, and fish cover; instead, these coefficients vary in direction.

Table 7: H1, Trans-boundary Groups

Trans-Boundary * Active Group	**Phosph. Level^ (-)**	**Nitrogen Level^ (-)**	**Turbidity Level^ (-)**	**Benthic MMI (+)**	**Riparian Cover+ (+)**	**Fish Cover+ (+)**
0 Year Lag	-0.086	0.207	0.267	3.143	-0.003	0.009
5 Year Lag	-0.074	0.182	0.237	7.775`	0.008	0.023
10 Year Lag	0.296	0.306	0.396	6.419	-0.103	0.007
Total Duration	-0.017	0.013	0.022	0.443	-0.001	0.002

^ variable log-transformed, + dependent variable square rooted
` $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Testing another aspect of inclusion and representation, Table 8 demonstrates the predicted change in environmental impact associated with the presence of a collaborative group when the collaborative group has a distinct technical advisory body as an accessory. Presumably, such groups expand the breadth of perspectives and knowledge a group can leverage, which in turn increases effectiveness. The binary indicator 'TECHNICAL BODY' reflects whether a given group has some form of technical advisory committee associated with it that is intended to provide expertise and input. Table 8 does not reveal a statistically significant relationship between the presence of a technical advisory body and group effectiveness; further, the direction of the estimated effects do not portend of a positive effect either. In fact, the consistent directionality for several metrics indicates that groups with a technical advisory body might be less effective; Table 8 predicts increases in phosphorus and turbidity levels, and decreases in benthic condition, riparian cover, and fish cover, for all model specifications. I return to these findings below, but briefly, it is possible that the presence of a technical advisory group evidences a group that is more deliberative than operational, and thus has less "on-the-ground' impact.

Table 8, H1, Technical Advisory Body

Technical Advisory Group * Active Group	Phosph. Level^ (-)	Nitrogen Level^ (-)	Turbidity Level^ (-)	Benthic MMI (+)	Riparian Cover⁺ (+)	Fish Cover⁺ (+)
0 Year Lag	0.423`	0.064	0.291	-4.632	-0.024	-0.019
5 Year Lag	0.204	-0.024	0.100	-4.315	-0.022	-0.035
10 Year Lag	0.569	-0.219	0.596	3.806	-0.071	-0.052
Total Duration	0.015	-0.022	0.008	-0.433	-0.004	-0.004

^ variable log-transformed, ^^ dependent variable square rooted
` $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The final aspect of Hypothesis 1 that I test is the level of group inclusiveness amongst local stakeholders (Table 9). The variable 'TOTAL TYPES' represents the sum of indicators for the inclusion of several different stakeholder types in the collaborative group (with local government actors and citizens representing the baseline for group membership). These stakeholder types were developed inductively through extensive review of group membership rosters. For each group, the TOTAL TYPES variable represents the sum total of binary indicators for the participation of tribes, business interests, Federal and state government agencies, agricultural interests, environmental NGOs (e.g., The Nature Conservancy), and universities or colleges in the collaborative group.

Table 9: H1, Stakeholder Types

Total Stakeholder Types	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	-0.079	0.039	0.009	0.429	0.003	-0.004
5 Year Lag	-0.067	0.071	0.011	0.030	-0.005	-0.007
10 Year Lag	-0.095	0.077	-0.086	0.797	0.004	-0.010
Total Duration	-0.005	0.007$^{\cdot}$	0.002	0.017	0.000	-0.001

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{\cdot}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9 also fails to evidence consistent, statistically significant evidence of increased group efficacy associated with increasing the diversity of stakeholders included in a group. In particular, the effects for phosphorus content and nitrogen content are in opposing directions, as are benthic community health and fish cover. While several estimates for nitrogen level are significant, these effects are positive (i.e., predict an increased level of nitrogen), which most definitely prevents me from rejecting the null hypothesis. I discuss the implications of these findings in more detail below.

H2: An increased level of collaborative watershed management group responsibility (whether group operates as an information sharing forum, a planning group, or a policy implementation body) results in better water quality outcomes.

The results in Table 10 indicate that groups tasked with planning and similarly intensive activities achieve greater environmental gains than do groups tasked only with coordination and less intensive work; further, groups tasked with management responsibilities (Table 11) such as implementation and enforcement appear to perform even better. However, none of these relationships are shown to be statistically significant.

Table 10: H2, Planning Responsibility

Planning * Group	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	-0.094	-0.193	-0.050	0.980	0.051	0.008
5 Year Lag	0.090	-0.101	0.233	-1.448	0.062	-0.001
10 Year Lag	0.470	0.206	0.416	1.934	0.117	-0.007
Total Duration	0.036	0.001	0.035	-0.034	0.004	0.000

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: H2, Management Responsibility

Management * Group	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	-0.619*	-0.280	-0.406	7.199	0.107$^{`}$	0.069
5 Year Lag	-0.507	-0.222	-0.284	8.916$^{`}$	0.128$^{`}$	0.046
10 Year Lag	-0.518	-0.209	-0.109	15.006*	0.124	0.091
Total Duration	0.004	0.000	0.011	0.379	0.007	0.002

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results for groups given management responsibilities are very consistent and strong in magnitude though, so while I am unable to reject the null hypothesis, it does appear that groups accorded greater responsibility achieve greater environmental gains. For instance, groups given management responsibility are estimated to decrease chemical content and turbidity relative to groups tasked with coordination or even planning, although these coefficients are not significant with the exception of phosphorus level for all active group (i.e., a zero year lag). Groups with management responsibility are also estimated to be much more effective at improving benthic community condition. A group that has management responsibilities and has been active at least five years is predicted to score 8.92 points higher on the benthic multi-metric index; groups active at least 10 years have almost twice as large an effect (15.01). The results associated with groups given planning responsibilities are insignificant and do not demonstrate any consistent patterns, indicating that the most marked difference is between groups given management responsibilities and those that are not.

H3: Increased formalization of a collaborative watershed group results in better water quality outcomes.

Similarly inconsistent results are found pertaining to hypothesis three (H3). Table 12 tests the predicted difference between groups that have a dedicated coordinator and those that do not. Across the six models and four specifications of an active group, only one coefficient is statistically significant (increased riparian cover for groups active at least 10 years). One notable finding however is that the predicted effects generally become stronger as the threshold for an active group is increased. This makes a good deal of sense, since the differences that might emerge between groups with and without a dedicated coordinator likely manifest over time and are less apparent on an immediate basis.

Table 12: H3, Dedicated Coordinator

Coordinator * Group	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	-0.063	0.089	-0.014	0.280	0.023	0.007
5 Year Lag	-0.272	0.099	-0.156	1.009	0.074	-0.011
10 Year Lag	-0.050	0.213	-0.016	4.106	0.171^{*}	0.028
Total Duration	-0.001	0.012	0.005	0.211	$0.008^{`}$	0.001

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Groups that are more institutionally formalized in terms of codification of group procedures, membership requirements, and other constitutional details in a bylaw document, are not shown to be statistically significantly different than groups that are not (Table 13). While more formalized groups are predicted to decrease phosphorus and turbidity levels and increase fish and (significantly) riparian cover, these same groups are also associated with lessened effectiveness in terms of nitrogen level reduction and improving the health of benthic communities.

Table 13: H3, Group Bylaws

Bylaws * Group	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	-0.248	0.182	-0.165	-1.957	0.066$^{`}$	0.009
5 Year Lag	-0.285	0.161	-0.401	-2.628	0.101*	0.018
10 Year Lag	-0.654	0.216	-0.097	-0.876	0.145*	0.104$^{`}$
Total Duration	-0.017	0.016	-0.017	-0.028	0.010*	0.004

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Further goal and objective formalization (Tables 14 and 15) do not show significant support for hypothesis three either. In fact, groups with more formalized goal sets are predicted to be less effective at improving riparian condition and phosphorus levels (with the negative effects on riparian being generally significant). The results for other metrics, including fish cover, benthic community health, and nitrogen level, are inconsistent and do not speak to a common theme.

Table 14: H3, Itemized Goals

Itemized Goals * Group	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	0.105	0.124	-0.135	0.693	-0.055	-0.012
5 Year Lag	0.161	0.110	-0.278	4.466	-0.080$^{`}$	-0.004
10 Year Lag	0.002	-0.216	-0.234	-4.467	-0.130*	0.061
Total Duration	0.009	-0.003	-0.018	-0.051	-0.008*	0.002

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: H3, Measurable Objectives

Measurable Objectives * Group	**Phosph. Level$^{\wedge}$ (-)**	**Nitrogen Level$^{\wedge}$ (-)**	**Turbidity Level$^{\wedge}$ (-)**	**Benthic MMI (+)**	**Riparian Cover^{+} (+)**	**Fish Cover^{+} (+)**
0 Year Lag	-0.057	0.199	-0.163	1.932	0.058	-0.053
5 Year Lag	0.180	0.279	-0.349	1.018	0.063	-0.070
10 Year Lag	0.016	-0.150	-0.152	4.384	0.080	-0.001
Total Duration	0.017	0.010	-0.020	-0.039	0.005	-0.004

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The most interesting takeaway from Tables 14 and 15 is that generally speaking, establishing specific goals or even measurable objectives does not appear to make collaborative watershed groups more effective. I return to this issue more fully in the discussion section below.

H4: Collaborative watershed groups in which local participants provide a larger degree of funding result in better water quality outcomes.

Contrary to hypothesis four, groups with higher levels of locally sourced funding appear to be less effective, not more effective (Table 16). Groups active for at least five years with at least 30% of total funds stemming from local sources increase the predicted effect of a group on stream nitrogen content by 37% (exp[0.317] = 1.37) for all active groups, and 58% (exp[0.455] = 1.58) for groups active at least five years. While no parameter for phosphorus level is statistically significant, each term is also positive and similarly strong in magnitude; for instance, groups active at least five years with a large proportion of local funding are associated with a 32% increase in the effect of a group on total phosphorus level (exp[0.278] = 1.32). Though the findings are not significant, locally funded groups are also predicted to achieve less gains in benthic condition and riparian cover, and only somewhat larger gains in in-stream fish cover.

Table 16, H4: Local Funding

(active group x >30% Local)	**Phosph. Level^ (-)**	**Nitrogen Level^ (-)**	**Turbidity Level^ (-)**	**Benthic MMI (+)**	**Riparian Cover+ (+)**	**Fish Cover+ (+)**
0 Year Lag	0.334	0.317`	0.078	-4.667	-0.015	0.007
5 Year Lag	0.278	0.455*	-0.146	-1.720	-0.041	0.005
10 Year Lag	0.061	0.072	-0.134	-3.460	-0.060	0.006
Total Duration	0.023	0.024	-0.011	-0.259	-0.003	0.002

^ variable log-transformed, ^^ dependent variable square rooted
` $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The relationship between group effectiveness and state funding, shown in Table 17, is inconsistent. State funding is predicted to significantly decrease a group's impact on nitrogen level (i.e., associated with higher levels of nitrogen), but (insignificantly) increase a group's impact of phosphorus content. All other effects are insignificant as well, though state funding is associated with a general enhancement of a group's impact on riparian cover and stream turbidity.

Table 17, H4: State Funding

(active group x >30% State)	Phosph. Level$^{\wedge}$ (-)	Nitrogen Level$^{\wedge}$ (-)	Turbidity Level$^{\wedge}$ (-)	Benthic MMI (+)	Riparian Cover^{+} (+)	Fish Cover^{+} (+)
0 Year Lag	-0.168	0.386*	-0.383	-1.839	0.006	-0.026
5 Year Lag	-0.276	0.571**	-0.279	-1.516	0.024	0.005
10 Year Lag	-0.697$^{`}$	0.079	-0.436	0.117	0.051	-0.025
Total Duration	-0.026	0.034*	-0.031	-0.054	0.003	0.000

$^{\wedge}$ variable log-transformed, $^{\wedge\wedge}$ dependent variable square rooted
$^{`}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, the results of Federal funding (Table 18) are somewhat inconsistent as well, in that Federal funding is significantly associated with groups active at least being much less effective at reducing nitrogen level (77% less effective for groups active at least five years (exp[0.571) = 1.77)), but more effective at reducing phosphorus levels and stream turbidity (though the effect of Federal funding on group effectiveness is only significantly related to decreased phosphorus levels across groups active at least ten years). While phosphorus level and nitrogen level are not highly correlated (they have 0.32 correlation coefficient in the sample), it is nonetheless curious that Federal funding affects each differentially. One potential reason is an unmodeled interaction between modes of agriculture and Federal funding. Agricultural sources generally account for more than 70% of phosphorus and nitrogen delivered into streams (Alexander et al. 2008), but nitrogen primarily comes from corn and soybean cultivation whereas phosphorus stems more equally from manure, urban runoff, corn and soybean cultivation and other crops (Alexander et al. [2008] find in the Mississippi River that the percentage breakdown is 37% 12%, 23%, and 18% respectively). Thus, the differential effects of Table 18 perhaps might stem from watershed in particular areas being more likely to be federally funded, or from federally funded programs being more focused on –or more able to achieve gains—on policies that affect sources of phosphorus and nitrogen differently.

Table 18, H4: Federal Funding

(active group x >30% Federal)	**Phosph. Level^ (-)**	**Nitrogen Level^ (-)**	**Turbidity Level^ (-)**	**Benthic MMI (+)**	**Riparian Cover+ (+)**	**Fish Cover+ (+)**
0 Year Lag	-0.168	0.386^{*}	-0.383	-1.839	0.006	-0.026
5 Year Lag	-0.276	0.571^{**}	-0.279	-1.516	0.024	0.005
10 Year Lag	$-0.697^{\prime}$	0.079	-0.436	0.117	0.051	-0.025
Total Duration	-0.026	$0.034^{\prime}$	-0.031	-0.054	0.003	0.000

^ variable log-transformed, ^^ dependent variable square rooted
$^{\prime}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Discussion

The results of this analysis provide some limited support for the conventional wisdom that effective collaborative groups include more stakeholders, fully engage in public policy and management issues, and have a stronger institutional presence (e.g., have a dedicated coordinator). However, several other conventions lack any support whatsoever in the above models. Technical advisory groups do not appear to enhance group impact, for instance. Likewise, while a lack of congruence between administrative and ecological boundaries is often espoused as a primary reason for engaging in collaborative environmental policy efforts, the lack of difference between groups that reach across state and provincial boundaries indicates that the benefits of such comprehensiveness might be limited to a more local level. While watershed condition is product of environmental behavior throughout a basin, it appears that that increasing small-scale comprehensiveness by involving more local stakeholders is more efficacious than increasing large-scale interaction involving numerous higher-level administrative bodies.

Another tension that emerges from these results is that while the presence of a collaborative group itself is strongly associated with water quality and watershed health (Table 4), this effect does not appear to vary considerably across the policy characteristics tested herein. In terms of the statistical analyses above, the presence of a watershed group is a significant source of variance in environmental outcomes, but several of the group characteristics tested fail to account for this variance. One potential reason is that the variables tested might not be the variables that drive group effectiveness. A notable omission, of course, are group funding levels. While one would presume that differential effectiveness associated with funding discrepancies is somewhat of a given, this relationship is worth testing to posit whether public agencies devoting funds to collaborative endeavors are getting any "bang" for their "buck." Unfortunately, funding data are not readily available for all groups included in this sample; for a second phase of this project, however, I am conducting a follow-up survey with contacts from each group identified in the data collection process. My intention is to use a brief form to both: (1) produce data that is not ubiquitously available via public documents, such as funding and participation levels; and (2) provide support for current data so as to conduct multimodal analysis and avoid unimodal biases that might arise in the current analysis (for instance, if some groups have a much stronger web presence than empirical presence). Of course, resource munificence alone cannot be the sole driver of group effectiveness. For instance, the findings above fail to identify a significant benefit associated with facilitating a technical advisory group, evidencing that policy makers should think carefully about what collaborative group funds are spent on.

These results also highlight the essential role of qualitative research in understanding the role and function of collaborative management. For instance, the extensive case studies conducted by Margerum (2011) speak to contextual variables and localized drivers of group efficacy that do not necessarily emerge in a larger-N cross-sectional analysis. In particular, while the representative sampling of the WSA/NRSA and grouping factors used in the multilevel model facilitate comparison between watersheds that "self-select" into the collaborative treatment group and those that do not, it is certainly possible that these same factors that motivate self selection result in environmental improvement without regard to the particulars.

Conclusion

It is easy to lose sight of the fact that collaborative governance also requires state resource expenditure, time and effort that could be applied elsewhere. This analysis has probed the "black box" of collaborative policy making and examined how specific group characteristics and design choices enhance (or detract from) the environmental impact associated with collaborative watershed management. The goal of this work is to help shift the decision framework from a binary approach in which the question is simply whether or not to engage in a collaborative governance approach to a more nuanced discussion in which policy makers can design and implement collaborative governance strategies to be maximally effective.

This analysis identifies a statistically significant difference between watersheds that are managed collaboratively and those that are not, but is unsuccessful in accounting for the source of that variation. This raises the possibility that collaborative watershed management might not have the effects it is often purported to have. If the organizational characteristics of a watershed group are unrelated to water quality improvements but the presence of a group is, then this indicates that collaborative watershed groups do not in and of themselves matter a great deal, as they are subject to self-selection. In other words, collaborative groups arise in context within which stakeholders are already motivated to improve and protect water quality, and the group is not a causal drive but instead an effect. This interpretation is not incompatible with the extant literature, which links successfully operating groups to the presence of social capital, trust, and norms of reciprocity amongst group members (e.g., Ansell and Gash 2008; Margerum 2011, Sabatier 2005). Stakeholders possessing considerable social capital and sharing a common interest in improving environmental conditions are presumably more likely to form and sustain a collaborative group and to foster environmental improvement; the favorable context drives each occurrence. This indicates that policy makers about whether collaborative groups are an efficient vehicle for environmental restoration and protection, or whether stakeholder interests and abilities (should they exist) are more efficiently channeled through other policy tools such as public-private partnerships or grants. Perhaps the central takeaway is that we as policy scholars and practitioners need to think more deeply about why we believe that collaborative groups are an effective vehicle for service delivery and how such delivery can be improved.

Going forward, there are two primary future directions for this research: (1) a survey conducted with group liaisons to develop a multimodal dataset and to produce contextual data and data that are not currently public (e.g., group expenditures); and (2) a continuation and expansion of the longitudinal analysis. The EPA is conducting a second NRSA over the course of 2013 and 2014, which will produce a third observation for the 357 watersheds sampled under both the WSA and first NRSA, and a two-period sample for larger rivers sampled under both the first and second NRSA. This will expand the breadth and depth of available data, which should provide more consistent model results and improve my ability to parse secular changes from effects attributable to the use of collaborative management. The spread of collaborative approaches in empirical practice has grown over the last few years, meaning that the “treatment” population will expand as the second NRSA data become available since more watersheds are managed collaboratively (and because groups will have been active longer in each watershed as well). The survey data will serve to better triangulate group characteristics, while the second NRSA will improve the statistical power of the analysis. Given current data, however, it appears unlikely that collaborative management has the environmental benefits it is generally purported to have.

References

Alexander, Richard B., Richard A. Smith, Gregory E. Schwarz, Elizabeth W. Boyer, Jacqueline V. Nolan, and John W. Brakebill. 2007. “Differences in Phosphorus and Nitrogen Delivery to The Gulf of Mexico from the Mississippi River Basin.” *Environmental Science & Technology* 42 (3) (December 21): 822–830. doi:10.1021/es0716103.

Altheide, David, Michael Coyle, Katie DeVriese, and Christopher Schneider. 2008. “Emergent Qualitative Document Analysis.” *Handbook of Emergent Methods*: 127–151.

Ansell, C., and A. Gash. 2008. “Collaborative Governance in Theory and Practice.” *Journal of Public Administration Research and Theory* 18 (4): 543–571.

Burby, R. J. 2003. “Making Plans That Matter: Citizen Involvement and Government Action.” *Journal of the American Planning Association* 69 (1): 33–49.

Carlson, C. 1999. “Convening.” In *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement, Ed. L. Susskind, S. McKearnon, and S. Carpenter*., 169–198. Thousand Oaks, CA: Sage.

Carr, G., G. Blöschl, and D. P. Loucks. 2012. “Evaluating Participation in Water Resource Management: A Review.” *Water Resources Research* 48 (11).

Coase, Ronald Harry. 1960. “Problem of Social Cost, The.” *Journal of Law and Economics* 3: 1–69.

Coglianese, C. 1999. “The Limits of Consensus: The Environmental Protection System in Transition: Toward a More Desirable Future.” *Environment: Science and Policy for Sustainable Development* 41 (3): 28–33.

Curtis, A., and I. Byron. 2002. *Understanding the Social Drivers of Catchment Management in the Wimmera Region*. Albury, NSW: Johnstone Centre for Research in Natural Resources and Society.

Dryzek, John S. 1997. *The Politics of the Earth: Environmental Discourses*. New York, NY: Oxford University Press.

Emerson, K., T. Nabatchi, and S. Balogh. 2012. "An Integrative Framework for Collaborative Governance." *Journal of Public Administration Research and Theory* 22 (1): 1–29.

EPA (U.S. Environmental Protection Agency). 2013. "National Rivers and Streams Assessment: 2008-2009." 2013. Washington, D.C.: U.S. Environmental Protection Agency. http://water.epa.gov/type/rsl/monitoring/riverssurvey/upload/NRSA0809_Report_Final_508Compliant_130228.pdf.

Gelman, A., and J. Hill. 2006. *Data Analysis Using Regression and Multilevel/hierarchical Models*. New York, NY: Cambridge University Press.

Gelman, Andrew. 2006. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do." *Technometrics* 48 (3): 432–435.

Guba, Egon G., and Yvonna S. Lincoln. 1994. "Competing Paradigms in Qualitative Research." *Handbook of Qualitative Research* 2: 163–194.

Hawkins, Charles P., Richard H. Norris, James N. Hogue, and Jack W. Feminella. 2000. "Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams." *Ecological Applications* 10 (5): 1456–1477.

Heikkila, T., and A. K Gerlak. 2005. "The Formation of Large-scale Collaborative Resource Management Institutions: Clarifying the Roles of Stakeholders, Science, and Institutions." *Policy Studies Journal* 33 (4): 583–612.

Huxham, Chris, and Siv Vangen. 2000. “Leadership in the Shaping and Implementation of Collaboration Agendas: How Things Happen in a (not Quite) Joined-up World.” *Academy of Management Journal* 43 (6): 1159–1175.

Imperial, M. T. 2005. “Using Collaboration as a Governance Strategy Lessons from Six Watershed Management Programs.” *Administration & Society* 37 (3): 281–320.

Innes, J. E, and D. E Booher. 1999. “Consensus Building and Complex Adaptive Systems.” *Journal of the American Planning Association* 65 (4): 412–423.

Karkkainen, B. C. 2002. “Collaborative Ecosystem Governance: Scale, Complexity, and Dynamism.” *Virginia Environmental Law Journal*. 21: 189–244. /z-afirst/.

Koontz, T. M, J. A Carmin, T. A Steelman, and C. W Thomas. 2004. *Collaborative Environmental Management: What Roles for Government?* Washington, D.C.: RFF Press.

Koontz, T. M, and C. W Thomas. 2006. “What Do We Know and Need to Know About the Environmental Outcomes of Collaborative Management?” *Public Administration Review* 66: 111–121.

Leach, W. D., C. M. Weible, S. R. Vince, S. N. Siddiki, and J. C. Calanni. 2013. “Fostering Learning through Collaboration: Knowledge Acquisition and Belief Change in Marine Aquaculture Partnerships.” *Journal of Public Administration Research and Theory* (5).

Lubell, M., A. D Henry, and M. McCoy. 2010. “Collaborative Institutions in an Ecology of Games.” *American Journal of Political Science* 54 (2): 287–300.

Margerum, R. D. 2008. “A Typology of Collaboration Efforts in Environmental Management.” *Environmental Management* 41 (4): 487–500.

———. 2011. *Beyond Consensus: Improving Collaborative Planning and Management*. Cambridge, MA: MIT Press.

Moore, Elizabeth, and Thomas Koontz. 2003. “Research Note a Typology of Collaborative Watershed Groups: Citizen-based, Agency-based, and Mixed Partnerships.” *Society &Natural Resources* 16 (5): 451–460.

Newig, J., and O. Fritsch. 2009. “Environmental Governance: Participatory, Multi-level–and Effective?” *Environmental Policy and Governance* 19 (3): 197–214.

Ostrom, Elinor. 2000. “Collective Action and the Evolution of Social Norms.” *The Journal of Economic Perspectives* 14 (3): 137–158.

Platt, Jennifer. 2006. “Evidence and Proof in Documentary Reserach: Part I, Some Specific Problems of Documentary Research.” In *Documentary Research*, edited by John Scott. Vol. 1. Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W, and A. S Bryk. 2001. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Vol. 1. New York, NY: Sage Publications.

Sabatier, P. A, W. Focht, M. Lubell, Z. Trachtenberg, A. Vedlitz, and M. Matlock. 2005. *Swimming Upstream: Collaborative Approaches to Watershed Management*. Cambridge, MA: MIT Press.

Sabatier, P. A, W. D Leach, M. Lubell, and N. W Pelkey. 2005. “Theoretical Frameworks Explaining Partnership Success.” *Swimming Upstream: Collaborative Approaches to Watershed Management*: 173–200.

Salamon, L. M. 2002. *The Tools of Government: A Guide to the New Governance*. Oxford University Press.

Smith, Graham. 2004. *Deliberative Democracy and the Environment*. New York: Routledge.

Susskind, L. E, J. Thomas-Lamar, and S. McKearnen. 1999. *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement*. Sage Publications, Incorporated.

Susskind, Lawrence. 1987. *Breaking the Impasse : Consensual Approaches to Resolving Public Disputes*. New York: Basic Books.

Wondolleck, J. M, and S. L Yaffee. 2000. *Making Collaboration Work: Lessons from Innovation in Natural Resource Management*. Island Press.

Yaffee, S. L, A. Phillips, I. C Frentz, P. W Hardy, S. Maleki, and B. Thorpe. 1996. *Ecosystem Management in the United States: An Assessment of Current Experience*. Island Press.

APPENDIX I: Example of Protocol Used to Code Watershed Groups

Note: I apply this process iteratively across available documents. Since many groups have distinct documents reference bylaws, membership, and funding, for instance, a group might initially receive a "0" for each category of stakeholder representation when coding the bylaws document; these variables will then be recoded when analyzing the membership roster.

Q1: Is this textual source an: (1) official group website; (2) annual group report; (3) group bylaw or charter document; (4) piece of authorizing legislation

If no → disregard
If yes → proceed to Question 2

Q2: Does the textual source contain language that addresses a group's purpose?

If no → proceed to Question 5
If yes → proceed to Question 3

Q3: Does the text speaking to a group's purpose present an itemized set of purposes?

If no → code Objective Formalization as "MISSION STATEMENT"
If yes → proceed to Question 4

Q4: Does the itemized set of purposes contain specific, measurable points of reference (e.g., "reduce total nitrogen level" instead of "improve water quality")

If no → code Objective Formalization as "GOALS"
If yes → code Objective Formalization as "MISSION STATEMENT"

Q5: Does the textual source contain language describing or listing group membership?

If no → proceed to Question 14
If yes → proceed to Question 6

Q6-Q13: Does text describing group membership list a tribe (or business, Federal agency, etc.) as a member of the group?

If no → code Tribal Representation as 0, proceed to next question
If yes → code "Tribal Representation" as 1, proceed to next question
---for Q6-12, proceed to next stakeholder type; for Q13, proceed to Q14 ---

Q14: Does the textual source detail group funding or budget data?

If no → proceed to next question 18
If yes → proceed to question 15

Q15-Q17: Do local/state/Federal funds constitute at least 30% of total group budget?

If no → code Local/State/Federal Funding as 0, proceed to next question
If yes → code Local/State/Federal Funding as 1, proceed to next question

Q18: Does textual source contain language describing membership requirements, voting procedures, and other constitutional details?

If no → code Bylaws as 0, proceed to Q19
If yes → code Bylaws as 1, proceed to Q19

Q19: Does textual source reference and describe a technical advisory body associated with the group?

If no → code Technical Body as 0, proceed to Q20
If yes → code Technical Body as 1, proceed to Q20

Q20: Does textual source contain specific reference to a group coordinator or facilitator?

If no → code COORDINATOR as 0, proceed to Q21
If yes → code COORDINATOR as 1, proceed to Q21

Q21: Does textual source identify year in which group was formed?

If no → proceed to Q22
If yes → code FORMATION YEAR as specified year

Q22-Q28: Does textual source contain language reference to group actions or responsibilities related to EDUCATION (e.g., group "runs environmental education programs in local schools")?

If no → proceed to next question
If yes → code GROUP ACTIVITY as "education"

Q23: Outreach (e.g., group "reaches out to local farmers")
Q24: Coordination (e.g., group "provides forum where agencies can share information")
Q25: Monitoring (e.g., group "conducts ongoing monitoring of stream pollutants")
Q26: Projects (e.g., group is "conducting restoration on Smith Creek near Auburn")
Q27: Planning (e.g., group is "charged with developing comprehensive action plan")
Q28: Management (e.g., group is "lead local entity for water improvement program")