International
Studies
Review

International Studies Review (2011)**13**, 579–605

ANALYTICAL ESSAYS: EVALUATION, SYNTHESIS, REFLECTION

The Effectiveness of International Environmental Regimes: Comparing and Contrasting Findings from Quantitative Research[1]

Helmut Breitmeier
Fern Universität in Hagen

Arild Underdal
University of Oslo

and

Oran R. Young
University of California, Santa Barbara

This article uses quantitative methods to deepen and broaden our understanding of the factors that determine the effectiveness of international regimes. To do so, we compare and contrast the findings resulting from two major projects: the Oslo-Seattle Project and the International Regimes Database Project. The evidence from these projects sheds considerable light on the determinants of regime effectiveness in the environmental realm. Clearly, regimes do make a difference. By combining models and data from the two projects, we are able to move beyond this general proposition to explore the significance of a number individual determinants of effectiveness, including the distribution of power, the roles of pushers and laggards, the effects of decision rules, the depth and density of regime rules, and the extent of knowledge of the relevant problem. We show how important insights emerge not only from the use of statistical procedures to *separate* the effects of individual variables but also from the application of alternative techniques, such as Qualitative Comparative Analysis (QCA), designed to identify *combinations* of factors that operate together to determine the effectiveness of regimes. We use our results to identify a number of opportunities for additional research featuring quantitative analyses of regime effectiveness. Our goal is not to displace traditional qualitative methods in this field of study. Rather, we seek to sharpen a set of quantitative tools that can be joined together with the extensive body of qualitative studies of environmental regimes to

[1]The authors thank Ivar Torgersen for assistance in data formatting, and the Department of Political Science, University of Oslo, for financial support. A preliminary version of this study was presented at the ISA 2009 Annual Convention. Useful and encouraging comments from the panel discussant, Peter M. Haas, and other participants are gratefully acknowledged. This version has benefited substantially from comments and suggestions by three anonymous reviewers. The data presented by the authors are provided in replication files at http://www.sv.uio.no/isv/english/people/aca/stvau1/index.html.

Breitmeier, Helmut, Arild Underdal, and Oran R. Young. (2011) The Effectiveness of International Environmental Regimes: Comparing and Contrasting Findings from Quantitative Research. *International Studies Review*, doi: 10.1111/j.1468-2486.2011.01045.x

strengthen our ability both to identify patterns in regime effectiveness and to explore the causal mechanisms that give rise to these patterns.

How effective are international environmental regimes in solving the problems that lead to their creation? Can we identify factors that determine the extent to which regimes are successful in these terms? These questions are easy to pose but difficult to answer, especially when we rely only on the largely qualitative research that constitutes the mainstream in studies of international regimes. In this article, we examine efforts to supplement the mainstream by using quantitative methods to address these questions. Our goals are to move beyond crude arguments about whether regimes matter, to evaluate the influence of specific determinants of effectiveness, and, in the process, to formulate recommendations regarding the next steps in this field of research.[2] In the process, we demonstrate the value of a research strategy that combines quantitative and qualitative methods to explore the determinants of effectiveness using several techniques of analysis at the same time.

Many qualitative studies of international regimes aim to shed light on the emergence and the performance of specific regimes through in-depth studies of individual cases. Others use several case studies to illustrate theoretical ideas about processes of regime formation (for example, the nature of institutional bargaining) or the determinants of regime effectiveness (for example, behavioral mechanisms through which regimes influence the actions of their members) (Haas 1992; Haas, Keohane, and Levy 1993; Young 1999; Parson 2003). These studies generally involve small numbers of cases; they seek to assess the extent to which different regimes conform to theoretical expectations by making qualitative comparisons across cases. This stream of qualitative research has produced important results; we know a number of things now that we did not know 25 years ago about the nature of regimes and the roles they play in guiding the course of international affairs (Young, King, and Schroeder 2008).

Nevertheless, research of this type has not allowed us to lay to rest broad and conflicting claims about the extent to which regimes matter at the international level. Recent years have brought a trickle of studies employing quantitative methods to assess the nature and role of international regimes. Some of these studies focus on specific concerns like compliance on the part of regime members with well-defined rules or the extent to which interactions between distinct regimes are synergistic or conflictual (Mitchell 2004, 2008; Oberthür and Gehring 2006). Two major projects have sought to create quantitative data sets relating to regimes more generally and to bring these tools to bear explicitly in efforts to measure and explain the effectiveness of international institutions. The results of one of these projects, known as the Oslo-Seattle Project, are reported in a volume entitled *Environmental Regime Effectiveness: Confronting Theory with Evidence* (ERE) (Miles, Underdal, Andresen, Wettestad, Skjærseth, and Carlin 2002). The initial findings of the second project, which draws on data available in the International Regimes Database (IRD), are reported in a volume entitled *Analyzing International Environmental Regimes: From Case Study to Database* (AIER) (Breitmeier, Young, and Zürn 2006; see also Breitmeier, Levy, Young, and Zürn 1996a,b. Copies of AIER come with a CD containing the entire IRD database) and in another book entitled *The Legitimacy of International Regimes* (Breitmeier 2008). Together, these studies provide some initial glimpses into the kinds of contributions that quantitative analyses can make to our understanding of how, when, and in what ways regimes matter.

[2]We place our study of regime effectiveness in the larger stream of research on social institutions and governance systems. See Peters (2005), Ostrom (1990).

In this article, we compare and contrast the findings of ERE and AIER in the interests of documenting what we know about the effectiveness of regimes, exploring differences in the literature regarding the determinants of effectiveness, and developing suggestions for next steps in research in this field. The next section briefly describes the approaches ERE and AIER adopt together with the procedures they use to generate results. The following section comments on areas in which the two studies yield common conclusions. The third section provides a more detailed analysis of areas in which the conclusions of ERE and AIER (appear to) differ and sometimes even conflict with one another. The final section draws on the preceding analysis to make recommendations for next steps in quantitative research on the effectiveness of international regimes.

ERE and AIER: Approaches and Procedures

Environmental regime effectiveness: confronting theory with evidence and AIER have a number of things in common that make it possible to compare and contrast their findings in a meaningful fashion. They also have differences that need to be noted before we move on to analyze these findings. First and arguably foremost, both projects are relatively conventional in the sense that they create data sets that include regimes in which the members are states and the constitutive provisions are set forth in legally binding agreements. ERE includes data pertaining to 14 international environmental regimes; the comparable figure for AIER is 23 regimes (see Table 1).[3]

While the initial universe of cases is composed of these discrete international environmental agreements, both ERE and AIER define the unit of analysis in such a way as to increase the number of cases available for quantitative analysis. ERE does this by breaking down regimes into components or phases before and after major transitions. AIER follows a similar path, defining elements to include distinct regime components as well as incorporating data on regimes before and after "watershed" changes. This provides ERE with a universe of 44 cases, 37 of which are usable in a study of regime effectiveness (Miles et al. 2002:434–435). For its part, AIER ends up with 92 regime elements. Since 80 regime elements included in AIER were coded independently by two experts and 12 by one expert, 172 regime elements are available for quantitative analyses of the effectiveness of regimes (Breitmeier et al. 2006:58–61).[4] Although these are relatively small universes of cases, they do open up a range of opportunities for quantitative analysis.

Both ERE and AIER generate quantitative data from "qualitative" case studies and expert assessments. ERE brought together a research team to develop case studies of the effectiveness of the regimes included and then used the results of these studies as a basis for coding a number of variables on ordinal scales. With regard to problem type, for instance, the ERE data protocol asks the coder to evaluate the perceptions of the parties and rank the results on a scale of 1–4 where 1 signifies that the parties regarded the problem as predominantly benign and 4 indicates a problem perceived as strongly malign. This procedure produces usable quantitative data on 33 variables.[5] The AIER makes use of "a common data protocol that identifies and designs a large set of variables relevant to all members of the universe of international regimes" (Breitmeier, Young, and Zürn 2007:44). In the development of the AIER, expert coders provided data in response to a protocol containing more than 200 variables.[6] This process

[3]Two of the ERE regimes (satellite communication and nuclear nonproliferation) treat problems that exist in issue areas other than the environment. Two of the 23 AIER regimes are bilateral in character.

[4]This difference calls for cautiousness in comparing estimates of statistical significance across the two data sets.

[5]The initial Oslo-Seattle codebook included 62 variables. For various reasons, the work reported in ERE relies only on data relating to 33 of these variables.

[6]In most cases, one coder brought a political science perspective and the other a legal perspective.

TABLE 1. International Regimes Included in the Oslo-Seattle and Analyzing International Environmental Regimes: From Case Study to Database (AIER) Projects

Oslo-Seattle Project (ERE)	*AIER*
• Dumping in the North Sea: The Case of the Oslo Commission	• Antarctic Regime
• Sea Dumping of Low-Level Radioactive Waste	• Baltic Sea Regime
• The Management of Tuna Fisheries in the West Central and Southwest Pacific	• Barents Sea Fisheries Regime
• The Vienna Convention and Montreal Protocol on Ozone-Layer Depletion	• Biodiversity Regime
• Cleaning Up the North Sea: The Case of Land-Based Pollution Control	• International Trade in Endangered Species
• The Convention on Long-Range Transboundary Air Pollution	• Climate Change Regime
• Satellite Telecommunication	• Danube River Protection
• The Management of High-Seas Salmon in the North Pacific	• Desertification Regime
• Nuclear Nonproliferation Regime	• Great Lakes Management Regime
• Mediterranean Action Plan	• Hazardous Waste Regime
• Oil Pollution from Ships at Sea	• IATTC Regime
• International Trade in Endangered Species	• ICCAT Regime
• International Whaling Commission	• International Regulation of Whaling
• Convention for the Conservation of Antarctic Marine Living Resources	• London Convention Regime
	• Long-Range Transboundary Air Pollution
	• North Sea Regime
	• Oil Pollution Regime
	• Protection of the Rhine Against Pollution
	• Ramsar Regime on Wetlands
	• Regime for Protection of the Black Sea
	• South Pacific Fisheries Forum Agency Regime
	• Stratospheric Ozone Regime
	• Tropical Timber Trade Regime

ERE, Environmental Regime Effectiveness: confronting theory with evidence.

produced the IRD, a large database that is accessible electronically and available for use by all interested researchers.

Whereas ERE focuses almost entirely on the question of effectiveness, AIER includes data on a wide range of themes relating to regimes more generally. Still, both projects treat effectiveness as a critical dependent variable, and they conceptualize this variable in a manner that is broadly comparable. ERE tracks effectiveness both in behavioral terms and in functional or problem-solving terms (Miles et al. 2002:4–7). It asks coders to specify changes in behavior relative to a hypothetical state of affairs absent the regime but otherwise the same as the real world (the no-regime counterfactual). ERE also asks coders to rate the performance of a regime on a continuum running from the no-regime counterfactual to an outcome defined as the collective optimum. The no-regime counterfactual takes on the character of a worst-case outcome against which actual achievements typically appear in a favorable light. The collective optimum is reached "when no further increase in benefits to one party can be obtained without leaving one or more prospective partners worse off" (Underdal 2002a: 9). The collective optimum sets a high standard against which actual achievements must be assessed.[7] AIER, by contrast, includes data on a broader set of consequences, encompassing information about effects framed as outputs, outcomes, and impacts as well as about effects outside the issue area of the regime in question. It has no direct analog of the concept of the collective optimum. The emphasis in both projects falls on the effects of regimes in fulfilling stated and unstated goals and in solving the problems that led to their creation. The data protocol for each project separates out the issue of causality, asking distinct questions not

[7]This standard is demanding also in terms of operationalization (see Young 2001; Hovi, Sprinz, and Arild Underdal 2003; Mitchell 2008).

only about the fulfillment of goals and the solution of problems but also about the causal force of the regime in bringing about these results.

At this juncture, ERE and AIER move in different directions. ERE proceeds to construct a model of regime effectiveness and to derive some hypotheses from this model that can be "tested" using data included in the project's data set (Miles et al. 2002:37 and 460–462). The ERE "core model" specifies that two complex variables called "problem malignancy" and "problem-solving capacity" account for the variance in "regime effectiveness" either directly or through their impact on an intervening variable designated "level of collaboration." It then proceeds to derive a series of hypotheses of the following sort: There is an inverse relationship between level of malignancy of the problem and the success of the regime in terms of problem solving. Several of the variables in the ERE core model are highly aggregated. Problem-solving capacity, for example, subsumes information about the institutional setting, the distribution of power, and the skill and energy of key players. Despite limitations imposed by its small universe of cases, ERE makes some effort to disaggregate these composite variables.

Analyzing International Environmental Regimes: from case study to database adopts a different strategy. This project focuses on a range of important issues in ongoing debates about international regimes and creates the IRD to allow analysts to test claims regarding these issues in quantitative terms (Breitmeier et al. 2007: 49–55). To illustrate, some analysts argue that international regimes cannot be effective because they lack enforcement mechanisms needed to induce subjects to comply with their requirements. AIER examines this proposition from a number of angles. It raises questions, for example, about what have become known as the enforcement and management models of compliance (Chayes and Chayes 1995; see also Downs, Rocke, and Barsoom 1996; Victor, Raustiala, and Skolnikoff 1998). Similarly, commentators often assert that international regimes cannot be effective because they rely on decision rules requiring consensus or even unanimous consent. AIER therefore uses its database to ask questions about the relative effectiveness of regimes employing different decision rules. The IRD also provides data usable to explore ideas about regime effectiveness that go beyond the sphere of regulation. It asks questions, for instance, about the roles regimes play in generating knowledge both about the nature of the problem and about the feasibility of different solutions.

These strategies lead ERE and AIER researchers to structure their arguments differently. Because ERE and AIER define their dependent variables in a manner that is reasonably similar, however, we can compare and contrast the results they produce, noting areas where these results are compatible and exploring areas where the results appear to differ. The differences between the projects with regard to research strategies constitute an advantage in some areas, since they allow analysts to engage in quantitative assessments of regime effectiveness that make use of two distinct modes of reasoning.

Common Findings

Both studies provide strong support for the proposition that regimes do matter, though the contributions they make vary considerably on the basis of a variety of conditions involving the nature of the problem, the character of the regime itself, and attributes of the broader setting in which it operates.[8] The two projects support the general conclusion that many regimes have a strong or at least a moderate causal effect in producing observed outcomes and impacts. A more detailed assessment of the ERE and AIER data sets demonstrates that this role

[8]These findings, we believe, should lay to rest at least the more extreme assertions of Strange (1983) and Mearsheimer (1994/1995).

has been caused by several regime attributes or by social practices triggered or reinforced by regimes. For example, both data sets show that regimes provided programmatic activities that helped to improve the knowledge base in the issue area or to reduce uncertainty about compliance on the part of regime members. As the authors of ERE observe, "[b]oth problem type and problem-solving capacity are statistically associated with both concepts of effectiveness and the direction of association is as we expected—negative for malignancy and uncertainty and positive for capacity" (Miles et al. 2002:443). Interestingly and somewhat at odds with initial expectations, the malignancy of a problem is not critical by itself as a determinant of effectiveness. But the combination of malignancy and uncertainty has a sharply negative effect in terms both of behavioral change and of movement in the direction of the collective optimum (Miles et al. 2002:442–443). The bottom line for the Oslo-Seattle research team is that regimes do make a "significant difference" (Miles et al. 2002:467).

The results that AIER produces are different in form but generally compatible with those emerging from ERE. Goal attainment is a relatively common occurrence with regard to the regimes included in the IRD; this is true even for cases that feature decision rules calling for consensus or unanimity (Breitmeier et al. 2006:124–131). The operation of the regime plays a large or at least a modest causal role in producing these results 35–75% of the time, depending on the nature of the decision rule in use. Not surprisingly, regimes are less effective in solving problems than in attaining goals. When problems improve considerably and regimes use consensus rules, for example, "… the causal impact of the regimes is very strong in 26.2% of the cases and significant in 44.3% of the cases" (Breitmeier et al. 2006:147). Somewhat similar results arise from AIER's analysis of the role of regimes in generating compliant behavior. Thus, "[i]n 52 of the 80 instances where states met or exceeded [regime] requirements, the regime had a large causal impact on state behavior" (Breitmeier et al. 2007:52).

Although the nature of the project makes it somewhat harder to tease out findings of a general nature about effectiveness, the overall message that AIER generates is that regimes frequently do matter; sometimes they matter a lot. Additionally, AIER sheds light on a variety of more specific issues of interest to regime analysts. To take one prominent example, AIER finds that "… neither the shallowness argument of Downs et al. (1996) nor the management approach of Chayes and Chayes (1995) can explain patterns of compliance with international environmental regimes" (Breitmeier et al. 2006:110). Compliance, AIER concludes, is a product of complex causality in which the most significant factors are incentives, institutional design, the rule of law, and the power of legitimacy (Breitmeier et al. 2006:110).

The two studies make it clear that pathways to effectiveness are complex, often involving a number of factors that interact with one another. As the authors of ERE observe, "… the odds for success measured as significant and major improvement in actor behavior are nineteen to one when a high capacity system deals with a non-malign problem that is fairly well understood, compared to one to eight when a low capacity system encounters a malign problem clouded in high uncertainty" (Underdal 2002b:458). A comparable finding from AIER deals with compliance and suggests that "… [a]dequate and even impressive rates of compliance with international environmental rules occur when appropriate incentive mechanisms are coupled with juridification, participation on the part of transnational NGOs in the rule-making process, and a responsive approach to the development of compliance mechanisms over time" (Breitmeier et al. 2006:112).

There is a sense in which ERE sets the bar regarding effectiveness higher than AIER does. ERE incorporates the idea of a collective optimum into its definition of effectiveness. It assesses behavioral change in relation to the hypothetical state

of affairs that would have existed in its absence. It is possible from this perspective for a regime to move a system some distance from the no-regime counterfactual yet receive a relatively low score on effectiveness because the outcome leaves a lot to be desired with regard to movement toward the collective optimum. Such situations cannot occur in the findings reported in AIER. The IRD measures effectiveness without reference to some notion of a collective optimum; it compares the state of goal attainment and problem solving that existed at the beginning and the end of a specific time period. A judgment about the causal role of a regime in accounting for observed changes with regard to goal attainment or problem solving is made only in another step following these measurements. It follows that a regime that gets a score of 4 or 5 in terms of both problem solving and causal influence in AIER's ranking system can end up with a lower score in the ERE ranking system. Once this difference is factored into our assessment, it becomes apparent that the general conclusions of the two studies regarding effectiveness are broadly compatible. Regimes do matter—sometimes significantly—but they ordinarily operate in circumstances where a number of interactive forces give rise to conditions of complex causality.

Differences and Extensions

We turn now to an examination of *differences* in the findings reported in the two studies. To set the stage properly, we distinguish between findings that are different yet *compatible* (at least as reported) and findings that are *incompatible* in the sense that they cannot both or all be true. When differences exist with regard to the questions asked, the causal models explored, the operational definitions of variables, or the data used in the analysis, divergences regarding specific findings would—by a strict interpretation—belong to the first category. From this perspective, few of the ERE and AIER findings are unambiguously incompatible. We have found only one that stands out as substantively important as well as clear-cut; it centers on the impact of decision rules. Yet, the two studies and the data sets they employ are sufficiently similar to warrant a critical examination of observations and conclusions that—although not incompatible as stand-alone statements—*diverge* in the sense that they invite interpretations that cannot be reconciled easily. Additional analyses using the two data sets may well reveal divergences that we have failed to detect.

We begin by examining descriptive statistics for the two data sets and move on to study causal relationships. Particularly for the latter purpose, we would have liked to take full advantage of the IRD, which is a richer database and which provided the basis for AIER. The use of a compare-and-contrast strategy, however, limits us to variables included in both data sets. As a consequence, we cannot test several interesting propositions from AIER against data from the Oslo-Seattle project. Nevertheless, combining the two data sets does allow us to identify new opportunities for exploring the sensitivity of some of AIER's conclusions to alternative specifications of multivariate models.

Some readers have come away from reading Miles et al. 2002 (ERE) and Breitmeier et al. 2006 (AIER) with the impression that AIER yields a more positive or upbeat result than ERE regarding the effectiveness of international regimes. What produces this impression, and what is its significance? At least three explanations seem plausible. One suggests that the AIER data set differs from the ERE data set with respect to the selection or coding of cases in such a way that overall regime achievements are "better" or that the task environments in which the regimes operate are more favorable. A second explanation focuses on the theoretical frameworks adopted and causal models examined in the two studies. In this regard, there is at least one potentially important difference. While ERE relies on a collective-action paradigm, AIER uses two approaches, the

collective-action framework and an alternative framework known as the "social-practice" paradigm (See also Young 2002b; Breitmeier 2008). Because the social-practice paradigm highlights non-coercive mechanisms, it may leave readers with a more upbeat understanding of institutions and political processes than those arising from the collective-action paradigm. A third explanation focuses on methodological issues. The main difference here is that ERE pushed its data set to the limit to make use of multivariate statistical techniques, while AIER relied essentially on (differentiated) bivariate analysis. There is no reason to assume that the former strategy would necessarily produce a gloomier picture, but in some instances, it might. One example may involve decision rules. In *bivariate* analysis, the two data sets give basically similar answers regarding the relevance of decision rules. It was only when ERE embarked upon a more ambitious *multivariate* analysis that differences emerged.

A First Cut: Descriptive Statistics

We can test the first of these explanations by using descriptive statistics. In Table 2, we have selected three main dimensions of regime achievement and two variables referring to features of the problems to be solved along with two variables describing factors included in the ERE concept of "problem-solving capacity." On all but one of these seven dimensions AIER's scores are "brighter." For three of the variables, the differences are hardly significant. In one of the other pairwise comparisons between problem change and problem-solving and arguably also in the pairwise comparison between compliance and behavioral change, a higher AIER score is to be expected due to ERE's more demanding standard.[9] For two dimensions (malignancy and power), however, scores differ significantly, even though the variables are fairly similar. We have therefore compared the scores on these two variables for the subset of regimes that are included in both data sets (and are not coded by the same person). With regard to problem malignancy, we found no systematic difference between the two data sets. But in the case of the distribution of power, there is more divergence. In about 70% of the cases where assessments diverge, AIER finds "pushers"[10] to be in a more

TABLE 2. The Samples as Described in the Two Databases—Descriptive Statistics for Selected Variables

	IRD		*ERE*	
Regime achievements (dependent variables)	Compliance	0.79	Behavioral change	0.66
	Problem change	0.63	Problem-solving	0.46
	Improved knowledge	0.78	Improved knowledge	0.82
Features of problem (independent variables) Capacity components (independent variables)	Problem understanding	0.68	Problem understanding	0.61
	Political malignancy	0.55	Political malignancy	0.72
	Majority decisions	18%	Majority decisions	12%
	Power advantage	70%	Power advantage	42%
	Pushers		Pushers	

(*Notes.* Some of the scales used in International Regimes Database [IRD] and Environmental regime effectiveness: confronting theory with evidence [ERE] differ. To facilitate comparison, we have—except for the two capacity components—translated the values originally assigned into scores on a *standardized* scale ranging from .00 [all cases assigned lowest value in the codebook] to 1.00 [all cases highest value].)

[9]One interesting concept developed to facilitate comparison across problem-regime complexes is that of "regime effort units" (see Mitchell 2004).

[10]In the international regimes literature, pushers are actors that become advocates or leaders in the formation and implementation of regimes.

advantageous position than ERE does.[11] Combining these observations does lead to the conclusion that AIER offers a somewhat "brighter" picture of regime performance and task environments than ERE does, though this finding seems to be attributable in part to differences in the selection of cases.[12]

Multivariate Analysis

More interesting than a simple comparison of descriptive statistics is the question of whether the two data sets point to the same determinants of "success" and the same causes of "failure." To answer this question, we need estimates of the effects of changes in one or more independent (and possibly also intervening) variables on regime effectiveness. Since we are dealing with small or at best moderate samples of cases, we need a research design that enables us to make efficient use of scarce data. To meet this requirement, we combine three different methodological techniques. One is partial correlation in which the statistical effect of a certain independent variable is measured stepwise by controlling for each of the other independent variables included in the model one by one (for example, through *trivariate* analysis). From the resultant computer runs, we report average partial correlations as well as the lowest and highest scores. Next, to be able to estimate effects of two or more variables simultaneously, we run binary logistic regressions. Here, we dichotomize each dependent variable and keep the number of values on all independent variables low (at most 3) in order to reduce the problem of empty cells. The first technique allows us to take advantage of information contained in nuances in the coding of each variable and to maximize the number of cases included. Logistic regression allows us to measure the effect of two or more variables simultaneously, although at the cost of sacrificing potentially important nuances in the original coding and in many instances reducing the number of cases included. We use both of these approaches primarily to *separate* the influence of individual variables. An equally important challenge is to identify *combinations* of factors that are necessary or sufficient to produce particular outcomes. For this purpose, we turn to a third technique of analysis, Ragin's (1987, 2000) qualitative comparative analysis or QCA.[13]

All these approaches have important limitations, especially when applied to small-N data sets. Nevertheless, we report results from the use of each of these techniques because we believe that even findings based on a small number of observations may provide interesting clues for interpreting observations and identifying hypotheses for further research. By combining the three techniques, we are able to get more out of the available data than we could by relying entirely on a single approach. Our analysis supports results that come out as clear and consistent across the two data sets using different methodological tools as well as alternative specifications of causal models. Results that differ significantly from one data set to the other or turn out to be highly sensitive to the specification of causal models call for more sophisticated analysis. Such differences may also indicate that more hard work is required to improve the validity and reliability of data sets available for regime analysis.

The Oslo-Seattle team conceived of variance in regime effectiveness as a function of two basic determinants—the nature of the *problem* and the *capacity* of regimes to solve or alleviate problems of the relevant types (Underdal 2002a). They argued that there are at least two factors that can make a governance

[11]ERE distinguished between power in the "basic game" (the system of activities to be governed) and power in the "negotiation game." The analysis in the book refers mainly to the former. The difference between ERE and AIER scores would have been less had we focused on the negotiation game.

[12]Separate analysis shows that the pattern is consistent across different "generations" of regimes.

[13]Manuals and software available at http://www.u.arizona.edu/~cragin/fsQCA.

problem hard to solve. The problem may be intellectually complex and demanding or for some other reason not well understood. It may also generate a politically "malignant" configuration of interests. ERE treats problem-solving capacity as a function of three main components—the institutional setting, the distribution of power, and the supply of informal leadership. This project's core model also includes one *intervening* variable—level of collaboration, a variable describing types of functions subject to collective (or "centralized") control. The ERE data set also includes information on some other regime properties and intermediate achievements that qualify as intervening variables.

We now confront ERE's core model with data from the two projects. We proceed in five main steps. First, we report trivariate partial correlations between each of the main independent variables and two variables describing outcomes. As a second step, we extend this analysis by exploring the impact of two sets of variables that may be considered intervening. A third step involves a switch to binary logistic regression, reporting results of four (in the case of ERE) or five to six (in the case of AIER) alternative models for each outcome variable. Our fourth step shows configurations of factors associated with high and low effectiveness. Finally, we turn to change *within* regimes, asking to what extent increasing or declining effectiveness can be accounted for by the same set of variables. The fact that the two data sets focus on somewhat different concepts of regime performance and, to some extent, on different independent and intervening variables calls for caution in comparing results.

With one remarkable exception, the overall impression arising from Tables 3 and 4 is one of similar effects. In both tables, regimes achieve lower performance scores when faced with problems that are poorly understood or characterized by severe political malignancy. Malignancy appears to be a somewhat more severe obstacle in the analysis of AIER data than in the ERE data set. Further scrutiny reveals that the impact of malignancy may also be contingent upon other factors. In ERE, Underdal (2002b:443) found that uncertainty and malignancy interact. A solid knowledge base serves to mute the effect of malignancy; high uncertainty and high malignancy interact synergistically and emerge as a "lethal" combination with regard to problem solving. In AIER data, a similar but weaker interaction effect is found for the contribution regimes make to problem change but not for their impacts on compliance. In most runs, regimes using majority voting do somewhat better than those requiring unanimity or consensus. This effect is somewhat stronger in the AIER data set.

TABLE 3. The Environmental Regime Effectiveness: Confronting Theory with Evidence Database—Regime Effectiveness

	Independent Variables				
Dependent Variables	*Uncertainty*	*Political Malignancy*	*Decision Rule in Use*	*{Institutional Capacity}*	*Distribution of Power*
Behavioral change					
Average	−.36	−.11	.01	.20	.40
Range	−.33* −.38*	−.02 −.09	−.11 .07	.18 .23	.38* .42*
Problem-solving					
Average	−.31	−.12	.08	.32	.21
Range	−.31 −.32	−.04 −.19	.00 .14	.27 .35*	.19 .23

(*Notes.* This table shows partial correlations between two measures of regime effectiveness and five independent variables when controlled for each of the other independent variables, in trivariate runs. [Institutional capacity is an index and is not used here as a control variable.] *$p<.05$, **$p<.01$, ***$p<.001$. $N = 29–35$. For definition of variables, see Appendix.)

TABLE 4. Analyzing International Environmental Regimes: From Case Study to Database—Regime Contribution to Compliance and Problem Change

	Independent Variables			
Dependent Variables	*Uncertainty*	*Political Malignancy*	*Decision Rule in Use*	*Distribution of Power*
Compliance, all members				
Average	−.30	−.27	.17	−.17
Range	−.23* −.34**	−.24* −.29*	.11 .21	−.16 −.18
Problem change				
Average	−.20	−.31	.10	−.15
Range	−.14 −.23*	−.26** .35***	.08 .13	−.10 −.20*

(*Notes.* See Table 3. $N = 75–129$).

For one of the independent variables (the distribution of power), however, we see a striking contrast. In the ERE data set, a power distribution in favor of pushers enhances regime effectiveness, particularly when measured as behavioral change. In our first cut at the AIER data, on the other hand, the corresponding correlations are consistently *negative*, although significant in only one of the runs. The difference detected in the coding of this variable would lead us to expect weaker impact in AIER data, but certainly not negative correlations.[14] But before concluding that we are faced with a major and unresolved disparity, we should consider also the results obtained through the use of regression and QCA tools.

The reasoning guiding the analysis in both studies included several regime properties that are or may be construed as *intervening* variables. In ERE, attention focused mainly on "level of collaboration." In order to facilitate comparison with AIER in this realm, we have added two other variables, referred to below as "selective incentives" and "regime contribution to the improvement of knowledge."[15] An intervening variable should be influenced by the independent variables included in the model and yet be sufficiently separate to have an identifiable effect on outcomes in its own right. All the variables selected meet this dual requirement.

Table 5 confirms that a high level of collaboration tends to enhance regime effectiveness. Also, regimes that contribute to strengthening the knowledge base tend to achieve higher scores, particularly regarding behavioral change. The presence of selective incentives does not seem to make much difference, at least in the case of problem-solving.

AIER does not explicitly designate particular factors as intervening variables. But the IRD does offer a rich menu of variables describing regime properties, and we may legitimately think of them as intervening. From this menu, we have selected four—referred to below as "depth and density of rules," "rules binding," "management approach," and "regime contribution to the improvement of knowledge." The latter is for all practical purposes a twin of its ERE counterpart. The first and second of these variables meet the correlational requirement of "intervening." But regime contribution to improving the knowledge base seems not much affected by any of the independent variables (except for its close relative, problem understanding).

[14]If a power advantage for pushers does enhance regime effectiveness, as hypothesized in ERE, the effect emerging from statistical analysis is likely to be weaker the less "demanding" the coding of this condition.

[15]The latter is not a regime property, but rather an intermediate result. Both studies conceive of regime contributions to the establishment of a solid base of consensual knowledge as an important step toward enhancing their performance as regulatory arrangements.

TABLE 5. Environmental regime effectiveness: confronting theory with evidence—The Impact of Intervening Variables

	Intervening Variables		
Dependent Variables	*Level of Collaboration*	*Selective Incentives*	*Improvement of Knowledge*
Behavioral change			
Average	.46	.13	.59
Range	.45* .47**	.10 .16	.59*** .60***
Problem-solving			
Average	.39	.02	.27
Range	.33 .45*	−.04 .08	.25 .29

(*Notes.* See Table 3. $N = 24–34$).

TABLE 6. Analyzing International Environmental Regimes: From Case Study to Database—The Impact of Intervening Variables

	Intervening Variables			
Dependent Variables	*Depth/Density of Rules*	*Rules Binding*	*Management Approach*	*Improvement of Knowledge*
Compliance, all members				
Average	.40	.05	.04	.28
Range	.36*** .46***	−.01 .13	−.07 .20	.22* .32*
Problem change				
Average	.33	−.27	−.08	.22
Range	.30*** .35***	−.22* −.36***	−.03 −.16	.14 .31**

(*Notes.* See Table 3. $N = 58–135$).

Table 6 indicates that regimes where rules qualify as "deep" or "dense" tend to produce significantly higher scores on both problem change and compliance.[16] As indicated by the narrow range of scores and high levels of statistical significance, this finding is robust. Interestingly, compliance is at best marginally higher—and regime contributions to problem change significantly *lower*—where rules are legally binding. A possible explanation of this result may be that regime members become more reluctant to undertake legally binding commitments as rules and regulations become more demanding (Downs et al. 1996). This interpretation would lead us to expect a negative correlation with depth/density of rules. But this expectation receives only weak support. Enforcement and management approaches do about equally well on both outcome dimensions. Regimes that contribute to improving the knowledge base tend to do better than regulatory arrangements, but the strength of that impact is weaker than in the Oslo-Seattle data set.

To estimate the effects of two or more variables simultaneously, we turn next to the use of logistic regression. For each of the outcome variables in the two studies, we present results for four (in the case of ERE) and five to six (in the case of AIER) alternative specifications of causal models. What is labeled Model 1 in the tables focuses on the four independent variables included in the ERE "core model," adapted so that the institutional capacity variable is replaced with decision rule in use. We then move on to add, stepwise, the variables that we have introduced above as "intervening." Since we are already stretching our data

[16]While rules may become deeper or denser as a regime matures, there are also variations among regimes in these terms from the outset.

TABLE 7. Environmental Regime Effectiveness: Confronting Theory with evidence (ERE)—Regime Effectiveness, Behavioral Change

Variables	*ERE-1*	*ERE-2*	*ERE-3*	*ERE-4*
Uncertainty	−1.642* (0.943)	−1.342 (0.879)		
Malignancy	−0.277 (0.508)			−0.924 (0.839)
Uncertainty + Malignancy			−0.598 (0.860)	
Dec. rule in use	0.716 (0.894)	0.819 (0.872)		
Power	0.727 (0.508)	1.059** (0.514)	0.915* (0.520)	
Inst. capacity			0.882 (0.913)	0.479 (1.137)
Level of collaboration		1.232 (1.091)		
Selective incentives			1.427 (0.900)	
Contribution to knowledge				2.050** (0.828)
Constant	1.034 (3.093)	−3.362 (3.205)	1.146 (2.344)	−2.240 (2.006)
Nagelkerke R^2	.345	.365	.402	.506
Corr. predictions (%)	77	76	76	82
N	35	37	37	27

(*Notes*. Logistic regression, binary. *$p<.10$, **$p<.05$, ***$p<.01$. Standard errors in parentheses.)

TABLE 8. Environmental Regime Effectiveness: Confronting Theory with Evidence (ERE)—Regime Effectiveness, Problem-solving

Variables	*ERE-5*	*ERE-6*	*ERE-7*	*ERE-8*
Uncertainty	−0.685 (0.951)			
Malignancy	−0.447 (0.467)			−1.510 (1.230)
Uncertainty + malignancy		−0.622 (0.919)	−0.999 (0.847)	
Dec. rule in use	1.151 (0.875)		1.150 (0.869)	
Power	0.645 (0.465)	0.330 (0.483)	0.543 (0.444)	
Inst. capacity		1.821** (0.931)		2.685* (1.426)
Selective incentives		0.014 (0.963)		
Contribution to knowledge				2.520* (1.352)
Constant	−1.563 (2.814)	−0.592 (2.496)	−1.161 (2.665)	−3.060 (2.512)
Nagelkerke R^2	.281	.351	.277	.653
Correct predictions (%)	77	78	78	77
N	31	32	32	26

(*Notes*. Logistic regression, binary. *$p<.10$, **$p<.05$, ***$p<.01$. Level of collaboration not included due to empty-cells problems. Standard errors in parentheses.)

sets to their limits, these models will include at most three of the independent variables from the ERE core model. In our analysis of the smaller data set, we replace one or more of these variables with an aggregate index. In our analysis of the AIER, we proceed by eliminating the type-of-problem variable that seems least important. Since we have identified the impact of decision rules and the distribution of power as critical issues, we keep these two variables in all models using IRD data.[17]

Tables 7 and 8 present results derived from the ERE data set. Broadly, the regression analysis corroborates the conclusions derived from the partial correlation analysis. A distribution of power favoring pushers emerges as the most important driver of behavioral change, while institutional capacity appears to be the key to effectiveness measured as problem-solving. Majority voting seems to

[17]We have, where feasible, run similar models with ERE data. Results are consistent with the findings summarized below.

have a positive though not statistically significant effect on both dimensions of effectiveness. To the extent that we have been able to compute results for the intervening variables, the overall picture corresponds well with that emerging from Table 5. But the impact of level of collaboration on behavioral change comes out as weaker than in the partial correlation analysis. The models predict (dichotomized) outcomes on the behavioral change dimension correctly in 77–89% of the cases. Corresponding figures for effectiveness measured as problem-solving range from 77% to 85%.

With one major and one minor exception, results from the analysis of IRD data correspond quite well to the preliminary conclusions arising from Tables 4 and 6. Looking first at correlates of compliance (Table 9), we see that a weak knowledge base emerges as the major hurdle, while coefficients for problem malignancy are much weaker. Majority voting has a positive effect in all models, but meets the $p<.10$ threshold only in one. Interestingly, all models including one or two of the regime properties we treat as intervening variables yield considerably higher scores on explained variance than the adapted ERE core model. This clearly indicates that regime properties are important determinants of effectiveness. Deep and dense rules consistently come out as a key to high compliance. Results for the management approach are unstable. Overall, the five models guide us toward correct predictions in 73–88% of the cases.

Looking at the contribution of regimes to problem change (Table 10), we note first that the overall "fit," as measured by the Nagelkerke R^2 as well as by prediction success, is significantly poorer than for compliance. The ERE core model yields correct predictions in only 63% of the cases. This is not impressive in dealing with dichotomized outcomes, where flipping a coin should yield correct predictions in about 50% of the cases for a large *N*. The key to effectiveness is once again deep and dense rules. Legally binding rules—and high compliance—are associated with small regime contributions to problem change. Both these findings are robust in the sense that they obtain also if we insert either variable into any of the other models included in Table 10. The competition between the enforcement and the management approaches remains undecided,

TABLE 9. Analyzing International Environmental Regimes: From Case Study to Database—Regime Contribution to Compliance

Variables	*IRD-1*	*IRD-2*	*IRD-3*	*IRD-4*	*IRD-5*
Uncertainty	−1.977*** (0.649)	−.2.165*** (0.800)	−2.186*** (0.743)	−3.463*** (0.977)	
Malignancy	−0.056 (0.427)	0.177 (0.534)			−0.265 (0.576)
Dec.rule in use	1.535* (0.881)	1.349 (1.026)	1.222 (1.042)	1.424 (1.211)	0.505 (1.067)
Power	0.133 (0.493)	0.612 (0.738)	0.768 (0.735)	0.516 (0.850)	1.312 (0.819)
Rules deep/dense		1.828*** (0.546)	1.902*** (0.546)	1.425** (0.630)	2.349*** (0.686)
Rules binding			−0.389 (0.731)		
Management approach				0.647 (1.708)	
Contribution to knowledge					0.518 (0.525)
Constant	1.1095 (2.085)	−3.670 (3.196)	−2.603 (3.134)	−2.027 (5.308)	−8.458** (3.824)
Nagelkerke R^2	.275	.537	.555	.608	.536
Correct predictions (%)	73	87	82	88	86
N	69	67	71	58	55

(*Notes.* Logistic regression, binary. *$p<.10$, **$p<.05$, ***$p<.01$).

TABLE 10. International Regimes Database (IRD)—Regime Contribution to Problem Change

Variables	*IRD-6*	*IRD-7*	*IRD-8*	*IRD-9*	*IRD-10*	*IRD-11*
Uncertainty	−0.395	−0.061				
	(0.465)	(0.514)				
Malignancy	−0.384	−0.576				
	(0.331)	(0.365)				
Uncertainty + malignancy			−1.002**	−1.048**	−0.865**	−1.511***
			(0.433)	(0.478)	(0.440)	(0.549)
Dec. rule in use	0.377	0.086	−0.620	−0.031	−0.597	−0.753
	(0.652)	(0.705)	(0.700)	(0.856)	(0.799)	(0.813)
Power	0.016	0.099	0.476	−0.232	0.210	0.542
	(0.354)	(0.401)	(0.483)	(0.481)	(0.459)	(0.607)
Rules deep/dense		0.853***	0.989***	0.698**	0.883***	1.443***
		(0.306)	(0.336)	(0.325)	(0.339)	(0.500)
Rules binding			−1.260**			−1.198**
			(0.510)			(0.585)
Management approach				−0.907		
				(1.302)		
Contribution knowledge					0.004 (0.361)	
Compliance						−1.125***
						(0.433)
Constant	0.711	−0.981	2.668	2.872	−0.018	5.276**
	(1.357)	(1.760)	(2.073)	(3.639)	(2.045)	(2.551)
Nagelkerke R^2	0.045	0.180	0.331	0.226	0.223	0.469
Correct predictions %	63	66	74	68	67	81
N	89	87	86	72	72	70

(*Notes.* Logistic regression, binary. *p<.10, **p<.05, ***p<.01.)

although most results tilt in favor of the former. The two problem features (uncertainty and malignancy) seem to interact synergistically, and the *combination* emerges as a significant obstacle to effectiveness (as in ERE).

Tables 9 and 10 do reveal one remarkable contrast with the findings reported from the partial correlation analysis. A distribution of power in favor of pushers now seems to have a weak *positive* effect on compliance and is insignificant for regime contributions to problem change. Ordinal regression analysis shows that the divergence between the two data sets in this regard is confined largely to the scores of neutrals and intermediates relative to those of pushers. There is one minor surprise as well. While majority voting is positively associated with compliance, results for regime contribution to alleviating the problem vary considerably, with more negative coefficients than positive.

So far, we have sought to *separate* effects of individual variables and determine how much of the observed variance in outcomes we can account for with different models. We now shift gears to search for *combinations* of factors associated with *particular* outcomes (for example, high compliance). For this purpose, we use the "crisp set" version of Ragin's QCA method. This approach is particularly useful in identifying factors or combinations of factors that are *sufficient* to bring about a particular outcome. In reporting results from this analysis, we make no claim that our findings provide general or foolproof recipes for success. Because the crisp set version of QCA requires dichotomous variables, our results are sensitive to the cutoff points used in distinguishing "high" and "low" scores as well as to specifications of the models examined. Some also are based on a very small number of observations. Still, we regard this type of analysis as important; it can

guide us toward the identification of causal pathways that are sufficiently promising to warrant further examination by researchers and serious attention on the part of practitioners.

We distinguish between pathways leading to *high* scores on measures of effectiveness and pathways associated with *low* scores. Recognizing that different types of problems may call for different "cures," we also distinguish between problems diagnosed as malignant and those coded as non-malignant. For each of the two data sets, we start with the adapted version of the ERE core model and move on to add, stepwise, the intervening variables examined in the statistical analysis.

Beginning with the ERE database, Table 11 shows clearly that there is more than one pathway to effectiveness. As expected, we find more pathways leading to behavioral change than to the more demanding goal of problem-solving, while the reverse is true for low scores. One pathway to high effectiveness does stand out from this analysis as a focus of attention. It includes a solid knowledge base, majority voting or high institutional capacity more broadly defined, and—for malignant problems—a distribution of power in favor of pushers. The most significant pattern associated with low effectiveness is less clear. But if we include runs with lower minimal requirements, we can conclude that a combination of high uncertainty and malignancy with a power advantage in favor of laggards and/or a demanding decision rule (or low institutional capacity) would be a fairly safe bet.

If we were to identify *one* particularly important key to effectiveness, Table 11 points to *knowledge*. A solid knowledge base is common to all but one of the pathways leading to high effectiveness, and a weak knowledge base occurs in six of the nine pathways associated with low effectiveness. A good understanding of the problem is not sufficient to guarantee success. But what makes knowledge uniquely important are the facts that only one of the high effectiveness pathways we have identified works without it and that it is important in dealing with malignant as well as non-malignant problems.

Results from the analysis of the AIER data set provide an interesting combination of support for and suggestions for revisions, extensions, and refinements of the conclusions emerging from ERE. The most unambiguous case of convergence concerns the importance of a good understanding of the problem to be solved. In Table 12, a solid knowledge base appears in *all* pathways leading to high compliance and positive problem change. Had we accepted lower minimal numbers of observations (3,3), we would have been able to identify pathways to high compliance also for the first three models, and a good understanding of the problem would have appeared in all of them. The two data sets thus converge on a crisp and clear message: although not a sufficient condition in itself, a solid knowledge base is an important ingredient in most recipes for success in creating international regimes. AIER's analysis using IRD data also reinforces what emerged as a more muted observation in the analysis of ERE data: knowledge seems more important on the "positive" side than on the "negative" side (there are many pathways to failure that do *not* involve high uncertainty).

The most important difference seems to be that the analysis using IRD data yields more mixed results for two of ERE's "capacity" variables: power and decision rules. In Table 12, both variables are somehow implicated in all of the pathways where they could be relevant, but rarely in a straightforward additive sense. This should not come as a big surprise. ERE's collective-action framework treats problem-solving capacity primarily as a matter of aggregating divergent preferences into collective decisions. An analysis of ERE data has already indicated that this formula does not work well for non-malignant problems (Underdal 2002b). Our analysis of AIER's use of IRD data supports this conclusion. With a higher proportion of non-malignant cases in the AIER data set, we would expect less prominent roles for power and decision rules.

TABLE 11. Environmental Regime Effectiveness: Confronting Theory with Evidence (ERE)—Pathways to High and Low Effectiveness

	High Effectiveness		*Low Effectiveness*	
Models	*Behavioral Change*	*Problem -solving*	*Behavioral Change*	*Problem-solving*
All valid cases				
ERE-QCA[1]	KNOW* MAJ*PUSH	ben*KNOW *MAJ*PUSH	–	UNC*MAL*CONS*lagg + UNC*MAL*lagg*cons
ERE-QCA[2]	BENKNOW *PUSH + PUSH* HICAP*	–	–	LOCAP*LAGG*malunc + LOCAP*lagg*MALUNC
N (range)	18–29	24–29	19–25	25–32
Malignant problems only				
ERE-QCA[1]	KNOW*MAJ*PUSH	KNOW*MAJ*PUSH	–	CONS*lagg
ERE-QCA[2]	KNOW*PUSH + KNOW*HICAP	KNOW*HICAP*PUSH	–	LOCAP*LAGG*unc + LOCAP*lagg*UNC
N	19–20	18–23	16–17	16–17
Non-malignant problems only				
ERE-QCA[1]	{KNOW*push + KNOW*MAJ}	{KNOW*maj*push}	{UNC*LAGG}	{UNC*LAGG}
ERE-QCA[2]	{KNOW*HICAP}	–	–	{UNC*LOCAP*LAGG}
N (range)	12	12	12	12

(*Notes.* Ragin QCA Crisp Set Solutions. Minimal requirements when all cases are included are 3 "right", 3 "false" observations. For the two subsets, the corresponding requirements are 2 and 2. Since this requirement would reduce the small subset of non-malignant cases to <10 valid observations, we report, in brackets, results obtained with 1 and 1. Model ERE-QCA[1] is the [adapted] ERE core model, including Problem understanding [*know/unc*], Malignancy [*mal/ben*], decision rule in use [*maj/cons*], and the distribution of power [*push/lagg*]. ERE-QCA[2] combines the two problem features [into *benknow/malunc*] and replaces decision rules with the broader concept of Institutional capacity [*hicap/locap*]. Upper case letters mean *presence* of a certain condition, lower case means *absence*, and – means that no pathway is found that meets the minimal requirements. * means "*and*", + means "*or*").

TABLE 12. Analyzing International Environmental Regimes: From Case Study to Database (AIER)—Pathways to Compliance and Problem Change

	High Effectiveness		*Low Effectiveness*	
Models	*Compliance*	*Problem Change*	*Compliance*	*Problem Change*
All valid cases				
IRD-QCA[1]	–	KNOW*BEN*maj*push	–	unc*mal*CONS*LAGG
IRD-QCA[2]	–	KNOW*BEN*maj*PUSH*DEEP	unc*mal*CONS*lagg*SHAL	–
IRD-QCA[3]	–	KNOW*BEN*maj*PUSH*DEEP*MAN	–	unc*mal*CONS*lagg*SHAL*MAN
IRD-QCA[4]	KNOW*BEN*deep*BIND*MAN	–	–	unc*mal*SHAL*NONBIND*MAN
N (range)	22–50	34–55	33–51	40–75
Malignant problems				
IRD-QCA[1]	–	–	#	–
IRD-QCA[2]	–	–	#	unc*CONS*lagg*SHAL
N (range)	10	11	<10	10
Non-malignant problems				
IRD-QCA[1]	KNOW*MAJ*push	–	UNC*CONS*lagg	–
IRD-QCA[2]	KNOW*MAJ*push*deep + KNOW*maj*push*DEEP	KNOW*maj*push	CONS*lagg*SHAL	unc*cons*lagg*SHAL
IRD-QCA[3]	KNOW*MAJ*push*deep*MAN + KNOW*maj*push*DEEP*MAN	KNOW*maj*push*deep*MAN	CONS*lagg*SHAL*MAN	–
IRD-QCA[4]	#	KNOW*deep*BIND*MAN	#	#
N (range)	24–34	13–45	28–35	37–47

(*Notes*. Ragin QCA Crisp Set Solutions. Minimal requirements when all cases are included: 5 "right", 5 "false" observations. For the two subsets, minimal requirements are 2 and 2. Model IRD-QCA[1] is the [adapted] environmental regime effectiveness: confronting theory with evidence [ERE] core model [see Table 11]. IRD-QCA[2] adds Rule depth and density [*deep/shal*], while model IRD-QCA[3] adds also Management approach [*man/nonman*]. In IRD-QCA[4], we have left out the two "capacity" variables of the ERE core model [power and decision rules] and added three AIER variables that we have treated as intervening [Rule depth and density, Rule binding [*bind/nonbind*], and Management approach]. Upper case indicates *presence* of a certain condition, lower case means *absence*, – indicates that no pathway is found that meets minimal requirements, and # indicates that there are too few cases left for the analysis. * means "*and*", + means "*or*").

TABLE 13. Intra-regime change

	ERE		*IRD*	
Regime Effectiveness	*Behavioral Change*	*Problem-solving*	*Compliance*	*Problem Change*
Increasing	72 [89]	55 [75]	80 [41]	82 [50]
Constant	75 [75]	63 [56]	29 [82]	22 [50]
Declining	– [0]	– [0]	– [0]	0 [0]
N	19	19	46	52

(*Notes*. Figures to the left in each cell show the proportion [in %] of *outcomes* that are correctly predicted, while figures in brackets show the proportion of *predictions* that fit outcomes. Predictions are based on the environmental regime effectiveness: confronting theory with evidence [ERE] core model "reinforced" by the intervening variable that emerged as the most important in Tables 4 and 6 ['level of collaboration' in the case of ERE, 'rules deep/dense' in the case of International Regimes Database (IRD)]. Our "predictions" assume that any change in *aggregate* score for this model will lead to a similar change in regime effectiveness. This is arguably an overly sensitive indicator, prone to predict more change than actually occurs. – indicates empty category.)

This poses an important question: what would a valid model for *non-malignant* problems look like? AIER took some initial steps toward answering this question, in part by adding the social-practice perspective and in part by including a wider range of independent and intervening variables. In Table 12, we present one model (IRD-QCA) in which power and decision rules are replaced by three regime properties—the depth and density of rules, the extent to which rules are legally binding, and the overall approach to compliance (enforcement/management). The results are encouraging, with a somewhat better overall fit than obtained for the ERE core model. More specifically, shallow rules are found in all negative pathways possible, while results for pathways leading to positive outcomes are mixed. A plausible interpretation—corroborated by ordinal regression analysis—is that avoiding a very low score on the depth/density variable is more important than achieving a top score. The binding rules variable appears in all pathways where it could appear and a positive value is associated with a positive outcome, but only for non-malignant problems that are well understood. Again, results regarding the management-versus-enforcement competition are inconclusive.

So far, we have not distinguished between variance *across* regimes and variance over time *within* regimes and regime components. But regimes are dynamic institutions that change continually (Young 2010). The ERE and AIER data sets offer limited opportunities for time series analysis. But both do record major transitions or "watersheds." We can use these distinctions to examine in a preliminary way how well the models we have explored above account for *intra*-regime changes in effectiveness scores.

In Table 13, we report results for the adapted ERE core model "reinforced" by the intervening variable that emerged as the most important in each data set.[18] The most striking observation is that this model does fairly well in predicting intra-regime increases in effectiveness from one period to the next (with a partial exception for increases in the most demanding standard of problem-solving) but fails completely in predicting *decline*. Not only does none of the predicted instances of decline materialize; the model also misses all instances of decline that we observe! Interestingly, in the analysis using IRD, data about two-thirds of the "errors" are overly optimistic predictions; for the Oslo-Seattle database, 75–80% of the "errors" are on the pessimistic side.

[18]This model corresponds to ERE-2 and IRD-2/IRD-7 in the logistic regression analysis.

Discussion

How should we interpret all these findings? In response to this question, we see five observations that are worthy of attention.

First, we have found that the AIER does offer a somewhat "brighter" characterization than ERE of regime performance as well as of task environments. Different samples of cases may account for most of this difference, but coding rules and practices also play a role.

Second, and more important, our analysis of the findings of the two projects yields conclusions that are largely similar or at least compatible regarding both conditions for "success" and causes of "failure." In multivariate analyses, to be more precise, the two projects yield basically *similar* results for *similar* variables, notably the two main problem features and at least one of the capacity components (decision rules). In both data sets, we also find a considerable amount of evidence indicating that programmatic activities (such as building a base of consensual knowledge and joint management of functions like monitoring and assessment) as well as the inclusion of certain regime properties (such as deep and dense rules) can become important tools for enhancing regime effectiveness over time (Breitmeier 2008: 87–89, 114–117). As noted above, a "reinforced" version of the ERE core model does fairly well using both data sets in accounting for intra-regime increases in effectiveness scores but fails to account for *declines*.

Third, we have made progress in resolving what appeared at first to be a major divergence concerning the role of power. The key to resolving this issue is to think of the role of power as *contingent* on the presence or absence of certain other factors. ERE concluded that the positive impact of pusher power is confined largely to malignant problems and, though less clearly, to effectiveness defined as behavioral change (Underdal 2002b: 449–451; 464). We find some support for the former conclusion also in AIER's results. The bivariate correlation between a power distribution in favor of pushers and regime contribution to compliance is 0.53 when interest incompatibility is at its highest and −0.14 when interests are largely convergent. The corresponding coefficients for regime contribution to problem change are 0.09 and −0.51*** respectively.[19] Moreover, our analysis of AIER's data set indicates that a power distribution in favor of pushers may enhance compliance but makes little difference regarding effectiveness defined as contribution to problem change. In another study based on the Oslo-Seattle database, Underdal (2008: 191) found that the impact of decision rules seems to depend on the distribution of power. More specifically, moving from consensus to qualified majority procedures appears to improve regime effectiveness primarily (perhaps only) where power is skewed in favor of pushers. A similar but weaker pattern is found in the AIER data set as far as regime contribution to problem change is concerned. Majority voting is not likely to lead to significant change unless pushers can form a winning coalition. What a less demanding decision rule can do is to help pushers translate a favorable configuration of power into formal regulatory actions. Our analysis also indicates that power and decision rules sometimes serve as functional equivalents. Both stringent decision rules and a power distribution in favor of laggards, for example, can contribute substantially to low effectiveness scores. With a high score on one of these factors, the other is likely to be redundant or have only a marginal impact.

Fourth, whatever the merits of these specific observations, the basic message emerging from our analysis is straightforward and clear. In measuring and explaining variance in the effectiveness of international regimes, we need more

[19]However, if we use incentives to defect as an indicator of problem malignancy, the pattern is reversed. *** indicates $p<.001$.

sophisticated models designed to allow us to capture conditional effects, synergies, and other types of interplay (See Young 2002a; Cash, Neil Adger, Berkes, Garden, Lebel, Olsson, Pritchard, and Young 2006; Underdal 2008; Young 2008). This observation echoes one of the main messages of the synthesis volume resulting from the work of the long-term project on the Institutional Dimensions of Global Environmental Change (IDGEC).[20] At least two different frontiers call for innovative contributions in future research. One involves supplementing research designed to *separate* the effects of individual variables with research focusing on *causal configurations*. We have taken a modest step in that direction using the crisp set version of QCA. The other frontier leads us from comparative statics to *dynamic* models of regime evolution.[21] Although it is beyond the scope of our current effort, we see this as an equally important challenge.[22] The two frontiers are related in the sense that dynamic models can help us to identify causal complexes and improve our understanding of how particular configurations of factors *co-produce* outcomes at different *stages* of the process.[23]

Finally, our analysis indicates that collective-action models have limited success in explaining (i) effectiveness defined in terms of problem change or problem-solving in contrast to less demanding results such as compliance or behavioral change and (ii) outcomes relating to problems characterized by low rather than high political malignancy. AIER took several promising steps in these realms by introducing a second paradigm (the social-practice approach) and by highlighting several design features and regime activities that seem important determinants of effectiveness. We have reported clear and robust conclusions pointing to the importance of factors such as a solid knowledge base and deep and dense rules. But much remains to be done to explore the details, to determine more precisely the merits of alternative frameworks, and to analyze how these factors interact with others. For these purposes, more sophisticated models that can help to enhance our understanding of causal *complexes* and process *dynamics* will constitute important steps forward.

Next Steps in Quantitative Research on Regime Effectiveness

Our comparative analysis of the ERE and AIER findings has helped to identify a number of methodological and theoretical issues requiring greater attention by those seeking to use quantitative data to shed light on the effectiveness of international regimes: (i) expanding the numbers and types of cases available for analysis; (ii) developing more sophisticated models to facilitate understanding of nonlinearities, conditional effects, and complex interplay among various factors, and (iii) exploring the roles of power and other alternative forces as determinants of regime effectiveness.

Expanding Numbers and Types of Cases

The results we have reported demonstrate the usefulness of the ERE and AIER data sets in testing hypotheses derived from different theoretical approaches to the study of regime effectiveness. Quantitative tools enable us to go some way toward meeting the epistemological challenge of confronting theories with empirical findings. But both the strength of specific hypotheses and the generalizability of theoretical propositions are inextricably linked to the number and

[20]A core project of the International Human Dimensions Programme on Global Environmental Change, IDGEC ran from 1998 through 2007. See Young et al. (2008).

[21]See the ideas presented in Axelrod (1997).

[22]For some initial steps in this direction, see Young (2010).

[23]Such configurations may well *shift* from one stage to another (Stokke 2010).

range of cases available for empirical testing. The development of the ERE data set and the IRD constitutes an important step forward in this realm. But more is needed.

What is the best way to move forward? One option would be to update case studies that have been coded already. Adding new regimes to the existing data sets is another option. The *UNEP Register of Treaties* includes a selection of 272 international environmental treaties and related instruments (UNEP 2005). But the real number of bi- and multilateral environmental agreements is much larger. To take full advantage of the lessons learned since the AIER and ERE data sets were created, however, we would have to expand the *range* of cases in three directions. First, we would want to study effects at the level of individual (member) states. Several other projects operate at this level of analysis. These studies typically focus on a single regime (for example, the regime addressing Long-Range Transboundary Air Pollution in Europe) or regime component (for example, LRTAP's 1985 Helsinki Protocol) (Helm and Sprinz 1999; Mitchell 2004; Ringquist and Kostadinova 2005). Data in this format provide opportunities for exploring the influence of domestic factors on state behavior as well as the impact of regimes on the capacity and behavior of member states. Second, useful new insights may be gained by adding regimes operating in different institutional settings. Thus, coding EU environmental directives and regulations would produce a data set that could help us assess whether the special character of the EU as a supra-national and highly legalized political system enhances compliance and problem-solving. Similarly, we believe that the study of regime effectiveness could benefit from including *transnational* governance systems, in part for what they accomplish on their own and partly for their contributions to the effectiveness of intergovernmental institutions (Hall and Biersteker 2003; Pattberg 2007; Delmas and Young 2009). Third, as Simmons demonstrates convincingly, policy domains differ in ways that may limit the scope of the validity of mainstream regime theory more than its pioneers anticipated (Simmons 2009). Models framed in terms of collective-action theory have a fairly strong record in domains such as environmental governance and international trade regulations, but they seem much less useful for understanding human rights regimes. Comparing and contrasting cases from different policy domains may help us to refine, differentiate, and perhaps also integrate models and propositions emerging from various subfields into more general theories of regime performance.

Developing More Sophisticated Models

Ecosystems are complex and dynamic. They frequently resist change until they cross a threshold or reach a tipping point that triggers nonlinear, frequently abrupt, and often irreversible changes. Any analysis of the effectiveness of regimes must confront the fact that the impacts of regimes may be hidden or obscured by the dynamics of the socio-ecological systems they seek to govern. Similar remarks are in order regarding the dynamics of social systems (Young 2002b). Governments may resist public demands for action regarding specific environmental issues (for example, climate change). But the start of a new administration can change resistance of this sort in short order. The formation of state interests responds to a range of factors that have little to do with the operation of individual regimes. Elites or particularly powerful interest groups may be able to block or veto decisions (Bachrach and Baratz 1962; Tsebelis 2002). Situations of this sort are often easy to explain in qualitative terms but may prove hard to incorporate into quantitative studies of regime performance. The nonlinear effects of social forces on the behavior of regime members may also reflect economic or cultural factors that are not considered adequately in the ERE and AIER data sets.

A strategy for improving our understanding of these complex and dynamic systems will have to include theoretical as well as methodological components. One important step could be to couple models of policy diffusion with models of cooperation. Policies can "co-evolve" through individual learning and adaptation as well as through joint decisions and deliberate coordination, and these processes often interact. Moreover, students of cooperation may learn from students of diffusion who have combined rationalist and constructivist concepts and models to analyze "norm cascades" and other nonlinear developments (Finnemore and Sikkink 1998). AIER took an important first step in that direction by supplementing the collective-action paradigm with a social-practice framework. Much remains to be done, however, before we are able to take full advantage of the complementary strengths of these and other approaches.

To understand complex systems, we also need to understand how different mechanisms and factors *co-produce* outcomes. As we have seen, QCA, even in its crisp set form, is a useful tool for identifying configurations of factors that lead to a certain outcome. "Fuzzy set" QCA may prove even more useful. However, the challenge is not primarily a matter of technical methods. More important are substantive questions about conditional effects and other types of interplay.[24] One lesson to be drawn from the analysis reported in this article is that trying to specify precisely the conditions under which a particular causal effect obtains can be a good place to start.

Finally, to study dynamics, we need time series data at the level of individual actors (for example, member states). AIER and ERE both provide information relating to regime development. But in both data sets, the format is too crude to get a good grasp on process dynamics.

Exploring the Role of Power and Other Drivers

If we were to pick one substantive issue calling for more refined analysis, the role of power would be a strong candidate. The significance of power construed either as the ability to get others to do things they would not otherwise have done or as control over outcomes has led regime analysts to pay attention to asymmetries in capabilities and resources that actors can bring to bear in pursuing their interests (Keohane and Nye 1977; Baldwin 1980:501). Two research projects completed in the early 1990s found evidence disconfirming the pure form of "hegemonic stability theory" regarding the creation and performance of regimes and demonstrated the limitations of conventional power-based explanations (Rittberger and Zürn 1990; Efinger, Mayer, and Schwarzer 1993; Young and Osherenko 1993; Hasenclever, Mayer, and Rittberger 1997). Still, the findings of the projects we consider in this article make it clear that power is a force to be reckoned with in efforts to understand the roles that regimes play in international affairs. The question is *how exactly* does this force work?

Our analysis indicates that the role of power is more complex than most of us anticipated. The effects of power are *contingent* on the presence or absence of certain other factors, such as the political malignancy of the problem, the decision rules in use, and the depth of knowledge of the problem. This article touches on such contingencies, but much remains to be done before we can claim to understand the interplay between power and other determinants of regime effectiveness.

Exploring this important topic further will require the collection of data pertaining to a broad spectrum of different forms of power conceptualized in a

[24]Interactions between or among distinct regimes is one type of interplay that has recently been studied in several projects. See for example, Underdal and Young (2004), Oberthür and Gehring (2006), Oberthür and Schram Stokke (2011).

more sophisticated manner. More specifically, we lack information about the strategies that states and other actors use in international venues to achieve desired outcomes. "Soft power," understood as the ability to achieve political goals through attraction rather than coercion, has become a relevant resource used by both the strong and the weak as they pursue individual and collective goals in world politics (Nye 2003). Intellectual, entrepreneurial, and structural leadership can influence both institutional bargaining and the effectiveness of regimes (Young 1991). States, non-state actors, expert groups, and individuals can achieve influence in discursive processes. In addition, the growing role of clubs or associations like the G-8 or the G-20 affects the context for conflict management within international regimes. Negotiations regarding issues like climate change have shifted in part to world economic summits. Powerful coalitions try to resolve important issues in political frameworks that operate outside the formal decision-making procedures of regimes. Further research on the role of power as a determinant of regime effectiveness must confront the challenge of taking these broader perspectives into account without incurring a crippling loss of rigor.

Concluding Remarks

Environmental regime effectiveness: confronting theory with evidence and AIER have significant limitations that we have sought to acknowledge explicitly in the course of our work. But, taken together, the two projects offer considerable encouragement regarding the contributions of quantitative analyses to understanding the effectiveness of international regimes. This is especially so when quantitative analyses, which are particularly useful in developing measures of association among variables, and qualitative analyses, which can help to explore the causal mechanisms underlying these relationships, are employed in tandem. The evidence we have analyzed from the ERE and AIER projects not only provides strong support for the proposition that regime matters, but it also allows us to begin to identify specific determinants of regime effectiveness that operate either individually or, more often, in combination with one another. Like any good scientific effort, this work also identifies new questions that call for additional research. There is no implication here that quantitative studies of regime effectiveness will displace the more familiar qualitative studies that constitute the mainstream of regime analysis. Rather, we advocate the use of a mixed strategy, deploying a toolkit that contains both qualitative and quantitative methods that can help us to make progress in understanding the factors leading to success in efforts to solve major problems through the creation of international regimes (Underdal and Young 2004).

References

Axelrod, Robert. (1997) *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press.

Bachrach, Peter, and Morton S. Baratz. (1962) Two Faces of Power. *American Political Science Review* 56: 947–952.

Baldwin, David. (1980) Interdependence and Power: A Conceptual Analysis. *International Organization* 34: 471–506.

Breitmeier, Helmut. (2008) *The Legitimacy of International Regimes*. Farnham, UK: Ashgate.

Breitmeier, Helmut, Marc A. Levy, Oran R. Young, and Michael Zürn. (1996a) *The International Regimes Database as a Tool for Studying International Cooperation. IIASA Working Paper WP-96-160*. Laxenburg, AT: International Institute for Applied Systems Analysis.

Breitmeier, Helmut, Marc A. Levy, Oran R. Young, and Michael Zürn. (1996b) *International Regimes Database (IRD): Data Protocol. IIASA Working Paper WP-96-154*. Laxenburg, AT: International Institute for Applied Systems Analysis.

Breitmeier, Helmut, Oran R. Young, and Michael Zürn. (2006) *Analyzing International Environmental Regimes: From Case Study to Database.* Cambridge, MA: The MIT Press.

Breitmeier, Helmut, Oran R. Young, and Michael Zürn. (2007) The International Regimes Database: Architecture, Key Findings, and Implications for the Study of Environmental Regimes. In *Politik und Umwelt, PVS Sonderheft 39/2007*, edited by K. Jacob, F. Biermann, P.-O. Busch, and P. H. Feindt. Wiesbaden: VS-Verlag, 41–59.

Cash, David W., W. Neil Adger, Fikret Berkes, Po Garden, Louis Lebel, Per Olsson, Lowell Pritchard, and Oran R. Young. (2006) Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World. *Ecology and Society* 11 (2): art. 8

Chayes, Abram, and Antonia H. Chayes. (1995) *The New Sovereignty: Compliance with International Regulatory Agreements.* Cambridge, MA: Harvard University Press.

Delmas, Magali, and Oran R. Young, Eds. (2009) *Governance for the Environment: New Perspectives.* Cambridge: Cambridge University Press.

Downs, George W., David M. Rocke, and Peter N. Barsoom. (1996) Is the Good News About Compliance Good News About Cooperation? *International Organization* 50: 379–406.

Efinger, Manfred, Peter Mayer, and Gudrun Schwarzer. (1993) Integrating and Contextualizing Hypotheses: Alternative Paths to Better Explanations of Regime Formation. In *Regime Theory in International Relations*, edited by V. Rittberger. Oxford: Oxford University Press.

Finnemore, Martha, and Kathryn Sikkink. (1998) International Norm Dynamics and Political Change. *International Organization* 52: 887–917.

Haas, Peter M. (1992) Epistemic Communities and International Policy Coordination. *International Organization* 46: 1–35.

Haas, Peter M., Robert O. Keohane, and Marc A. Levy, Eds. (1993) *Institutions for the Earth: Sources of Effective International Environmental Protection.* Cambridge, MA: The MIT Press.

Hall, Rodney B., and Thomas J. Biersteker. (2003) The Emergence of Private Authority in the International System. In *The Emergence of Private Authority in Global Governance*, edited by R. B. Hall, and T. J. Biersteker. Cambridge, UK: Cambridge University Press.

Hasenclever, Andreas, Peter Mayer, and Volker Rittberger. (1997) *Theories of International Regimes.* Cambridge, UK: Cambridge University Press.

Helm, Carsten, and Detlef Sprinz. (1999) *Measuring the Effectiveness of International Environmental Regimes. PIK Report No. 52.* Potsdam: Potsdam Institute for Climate Impact Research.

Hovi, Jon, Detlef F. Sprinz, and Arild Underdal. (2003) The Oslo Potsdam Solution to Measuring Regime Effectiveness. *Global Environmental Politics* 3: 74–96.

Keohane, Robert O, and Joseph S. Nye. (1977) *Power and Interdependence: World Politics in Transition.* Boston: Little Brown.

Mearsheimer, John J. (1994/1995) The False Promise of International Institutions. *International Security* 19: 5–49.

Miles, Edward L., Arild Underdal, Steinar Andresen, Jørgen Wettestad, Jon B. Skjærseth, and Elaine M. Carlin. (2002) *Environmental Regime Effectiveness: Confronting Theory with Evidence.* Cambridge, MA: The MIT Press.

Mitchell, Ronald B. (2004) Quantitative Analysis in International Environmental Politics: Toward a Theory of Relative Effectiveness. In *Regime Consequences: Methodological Challenges and Regime Consequences*, edited by A. Underdal, and O. R. Young. Dordrecht: Kluwer Academic Publishers.

Mitchell, Ronald B. (2008) Evaluating the Performance of International Institutions: What to Evaluate and How to Evaluate It. In *Institutions and Environmental Change: Principal Findings, Applications, and Research Frontiers*, edited by O. R. Young, L. A. King, and H. Schroeder. Cambridge, MA: The MIT Press, pp. 79–114.

Nye, Joseph S. (2003) *The Paradox of American Power: Why the World's Only Superpower Can't Go It Alone.* New York: Oxford University Press.

Oberthür, Sebastian, and Thomas Gehring, Eds. (2006) *Institutional Interaction: How to Prevent Conflicts and Enhance Synergies between International and European Environmental Institutions.* Cambridge, MA: MIT Press.

Oberthür, Sebastian, and Olav Schram Stokke, Eds. (2011) *Managing Institutional Complexity: Regime Interplay and Global Environmental Change.* Cambridge, MA: The MIT Press.

Ostrom, Elinor. (1990) *Governing the Commons: The Evolution of Institutions for Collective Action.* Oxford: Cambridge University Press.

Parson, Edward A. (2003) *Protecting the Ozone Layer: Science and Strategy.* New York: Oxford University Press.

Pattberg, Phillip. (2007) *Private Institutions and Global Governance: The New Politics of Environmental Sustainability*. Cheltenham, UK and Northampton, MA: Edward Elgar Publishing Ltd.

Peters, B. Guy. (2005) *Institutional Theory in Political Science: The New Institutionalism*. London: Sage Publications.

Ragin, Charles C. (1987) *The Comparative Method*. Berkeley: University of California Press.

Ragin, Charles C. (2000) *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Ringquist, Evan J., and Tatiana Kostadinova. (2005) Evaluating the Effectiveness of International Environmental Agreements: The Case of the 1985 Helsinki Protocol. *American Journal of Political Science* 49: 86–102.

Rittberger, Volker, and Michael Zürn. (1990) Towards Regulated Anarchy in East-West Relations: Causes and Consequences of East-West Regimes. In *International Regimes in East-West Politics*, edited by V. Rittberger. London: Pinter.

Simmons, Beth A. (2009) *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press.

Stokke, Olav S. (2010) *A Disaggregate Approach to International Regime Effectiveness: The Case of Barents Sea Fisheries*. Oslo: University of Oslo, Department of Political Science (Dr. Philos. Dissertation).

Strange, Susan. (1983) Cave! Hic Dragones: A Critique of Regime Analysis. In *International Regimes*, edited by S.D. Krasner. Ithaca, NY: Cornell University Press.

Tsebelis, George. (2002) *Veto Players: How Political Institutions Work*. New York: Russell Sage Foundation.

UN Environment Program. (2005) *Register of Treaties and Other Agreements in the Field of the Environment*. Nairobi: UNEP.

Underdal, Arild. (2002a) One Question, Two Answers. In *Environmental Regime Effectiveness: Confronting Theory with Evidence*, edited by Edward L. Miles, Edward L. Underdal, Arild Andresen, Steinar Wettestad, Jorgen Skjaerseth, Jon Birger Carlin, and Elaine M. Carlin. Cambridge, MA: The MIT Press.

Underdal, Arild. (2002b) Conclusions: Patterns of Regime Effectiveness. In *Environmental Regime Effectiveness: Confronting Theory with Evidence*, edited by E.L. Miles. Cambridge, MA: The MIT Press.

Underdal, Arild. (2008) Determining the Causal Significance of Institutions: Accomplishments and Challenges. In *Institutions and Environmental Change: Principal Findings, Applications, and Research Frontiers*, edited by O.R. Young, L.A. King, and H. Schroeder. Cambridge, MA: The MIT Press.

Underdal, Arild, and Oran R. Young, Eds. (2004) *Regime Consequences: Methodological Challenges and Research Strategies*. Dordrecht, NL: Kluwer Academic Publishers.

Victor, David G., Kal Raustiala, and Eugene B. Skolnikoff, Eds. (1998) *The Implementation and Effectiveness of International Environmental Commitments: Theory and Practice*. Cambridge, MA: The MIT Press.

Young, Oran R. (1991) Political Leadership and Regime Formation: On the Development of Institutions in International Society. *International Organization* 45: 281–308.

Young, Oran R., Ed. (1999) *The Effectiveness of International Environmental Regimes: Causal Connections and Behavioral Mechanisms*. Cambridge, MA: The MIT Press.

Young, Oran R. (2001) Inferences and Indices: Evaluating the Effectiveness of International Environmental Regimes. *Global Environmental Politics* 1: 99–121.

Young, Oran R. (2002a) Are Institutions Intervening Variables or Basic Causal Forces?: Causal Clusters vs. Causal Chains in International Society. In *Millennium Reflections on International Studies*, edited by M. Brecher, and F. Harvey. Ann Arbor: University of Michigan Press.

Young, Oran R. (2002b) *The Institutional Dimensions of Environmental Change: Fit, Interplay, and Scale*. Cambridge, MA: The MIT Press.

Young, Oran R. (2008) Building Regimes for Socioecological Systems: Institutional Diagnostics. In *Institutions and Environmental Change: Principal Findings, Applications, and Research Frontiers*, edited by O.R. Young, L.A. King, and H. Schroeder. Cambridge, MA: The MIT Press.

Young, Oran R. (2010) *Institutional Dynamics: Emergent Patterns in International Environmental Governance*. Cambridge, MA: The MIT Press.

Young, Oran R., and Gail Osherenko. (1993) Testing Theories of Regime Formation: Findings from a Large Collaborative Research Project. In *Regime Theory and International Relations*, edited by Volker Rittberger. Oxford: Oxford University Press, pp. 223–251.

Young, Oran R., Leslie A. King, and Heike Schroeder, Eds. (2008) *Institutions and Environmental Change: Principal Findings, Applications, and Research Frontiers*. Cambridge, MA: The MIT Press.

Technical Appendix: Transformation of Variables and Coding Procedures[25]

The International Regimes Database (used in AIER)

Political Malignancy
RF7, variable 101G (*INTEREST_DISOBEY*)
+ RF9, variable 101I (*INTEREST_INCOMPATIBILITY*)

Distribution of Power
RF19, variable 102C (*POWER_SETTING_SYMMETRY*), combined with background information in Part I. Scores assigned specifically for the purposes of this analysis.

Compliance
RC5, variable 303A (*CONFORMITY ALL_MEMBERS* x *CONFORMITY_CAUSAL*)

Problem Change
RC11, variable 304A (*PROBLEM_CHANGE* x *PROBLEM_CHANGE_CAUSAL*)

The Oslo-Seattle Database (used in ERE)

Political Malignancy
Var15 (*Type of Problem*)
If Var15 ≥ 3, then add
Var18 (*Symmetry*)
+ Var28 (*Cumulative Cleavages*)

Institutional Capacity
Var40a (*Decision Rule in Use*)
+ Var43 (*Fast Track Options*)
+ Var44 (*Role of Secretariat*)
+ Var45 (*Role of Conference Presidents and Committee Chairs*)

Where a transformation produces "too many" values on the new variable (leading to severe empty-cells problems), we have merged values. For the QCA analysis, "intermediate" values are left out of "high" and "low" categories.

[25]The data files used for the analysis and details about recoding and transformation of variables will be made available on a website.