

Article

Humans and the Water Environment: The Need for Coordinated Data Collection

John B. Braden ^{1,*}, Maria Christina Jolejole-Foreman ¹ and Daniel W. Schneider ²

¹ Department of Agricultural & Consumer Economics, University of Illinois, 1301 W. Gregory Drive, Room 326, Urbana, IL 61801, USA; E-Mail: jolejoll@illinois.edu (M.C.J.-F.)

² Department of Urban & Regional Planning and Illinois Natural History Survey, University of Illinois, 611 Lorado Taft Drive, Room 111, Champaign, IL 61820, USA; E-Mail: ddws@illinois.edu

* Author to whom correspondence should be addressed; E-Mail: jbb@illinois.edu; Tel.: +1-217-333-5501; Fax: +1-217-333-5538.

Received: 21 October 2013; in revised form: 25 November 2013 / Accepted: 27 November 2013 / Published: 19 December 2013

Abstract: Efforts to observe humans in relation to nature over time and at large scale are few and disjointed in ways that impede progress in building scientific foundations for sustainability. Two water-oriented national-scale case studies highlight the challenges of integrating existing natural system and social system data: one concerns the influence of environmental attitudes and water quality on water conservation efforts; the other explores relationships between environmental attitudes, water quality and recreation behavior. The case studies show that coupled research conducted at large scale can yield new insights, but uncoordinated data limit meaningful inference. We propose salient features of a coordinated observation program for water.

Keywords: environmental behavior; sustainability science; science policy; survey research; water conservation; water recreation

1. Introduction

This paper provides evidence that fundamentally new insights about human-water relationships are possible from coupled observation and analysis at large scale, but current data programs are insufficiently coordinated to provide generalizable results. These findings motivate a proposal to advance science policy toward data collection designed specifically for coupled observation across domains.

“Water problems” are fundamentally human problems [1,2]. People want water in greater or lesser quantities or in different locations than nature supplies, their actions degrade and divert water and aquatic ecosystems, and their needs change over time with population growth and migration [3,4], technology, and climate [5]. Sustainable clean water supplies and protection of water resources are priorities domestically [6] and internationally [7]. Much water infrastructure is due for costly replacement [8]. Solutions will involve changes in institutions, technologies, behaviors and the natural environment [9] and will evolve through time and circumstance [10]. Understanding human perceptions, motivations and behaviors, both individually and collectively, and their interactions with context is critical to addressing water challenges.

Theories of human-nature interactions should encompass diverse demographic, ecological, economic, institutional and temporal environments. One-off studies undertaken at local or watershed scale, while often important in their own right, may not span relevant gradients of variation, thereby limiting their applicability elsewhere. For example, social adaptation to drought may differ in regions where drought is frequent and people, infrastructure, and institutions are already attuned to it compared to regions where it is rare and response mechanisms are infrequently tested. Moreover, the study of drought response requires inter-temporal comparison of drought and non-drought conditions. Systematic and repeated observation at large spatial and temporal scale can support science that integrates across natural and social contexts to enable rigorous testing of fundamental relationships [1].

The primary scientific alternative to large-scale, statistically-designed observation is meta-analysis of local studies (e.g., [11–13]). Meta-analyses are inherently limited by the availability of studies. Those studies are often selected for their uniqueness and rarely are representative of the population of potential studies. Findings are unlikely to support generalization [14,15]. Thus, meta-analyses complement rather than substitute for studies based on large-scale, statistically-designed observation.

Various US programs collect water-related data, but the absence of coordination impedes coupled scientific research [16,17]. Rarely are the programs conducted with joint objectives or common sampling frames. The data are gathered at different temporal and spatial scales and levels of processing. Some use randomized sampling while others are opportunistic or driven by regulation. Some are precisely geo-referenced while others—particularly human data for reasons of privacy—are aggregated and averaged to postal code, census or jurisdiction scale. Data sets that ostensibly should relate to one another may have been gathered asynchronously with different sampling frames and may not reflect the same phenomena.

This paper presents concrete illustrations of the promise of human-water observation at large scale and the limitations of existing data collection efforts toward that end. We focus on two overarching questions: (OQ1) Do the best available large-scale social and natural system data sets about water adequately support testing of fundamental coupled research questions? (OQ2) Does coupled analysis at large scale enable new insights about human-water interactions? To provide evidence on these

questions, we present two case studies to illustrate fundamental research questions that require coupled data: Case Study 1 (CS1) explores two questions: (RQ1) Do scientifically-measured water quality parameters influence attitudes about water quality and are these relationships stable geographically? (RQ2) Do measured water quality and environmental attitudes influence water conservation efforts and are these relationships stable geographically? Case Study 2 (CS2) explores a third research question: (RQ3) Does water quality affect participation in recreation and is this relationship stable geographically? Answers to these questions would help in the design of programs aiming to influence water use and recreational behaviors. No previous studies have considered these questions at national scale.

2. Data Congruence

We explore these questions using representative national scale US data, which approximates continental scale. This scale encompasses significant spatial variation in climate, ecology, economic activity, institutions, and demographic characteristics. At the time of this study, 2011–2012, the best available, statistically-designed and pertinent national-scale data sets were the US Environmental Protection Agency's (USEPA) 2004–2005 Wadeable Streams Assessment (WSA) [18] for water quality measures, the National Opinion Research Center's (NORC) [19] environmental module of the 2010 General Social Survey (GSS) for attitudinal and self-reported water conservation efforts and the U.S. Forest Service's 2009 round of the National Survey on Recreation and the Environment (NSRE) [20] for water-related outdoor recreation activities. The WSA data are collected for specific geographic points using a sampling procedure designed to be representative at ecoregional [21] and national scales. Sites are assigned weights (as stream distances) based on their probability of inclusion in the random sampling frame. Using these weights, statistical inferences about different spatial or subpopulations can be generated as long as there are sufficient members of the subpopulation [22]. The GSS and NSRE data are based on responses obtained from individual households, but they are available only at polygon scale; census tracts for GSS and zip codes for NSRE are the smallest geographic units. On average, census tracts contain approximately 4000 people and zip code units contain approximately 7500 people. The original household data are upscaled and averaged to protect respondent identities. The GSS and NSRE sampling frames are statistically representative at national scale. The geographic distributions of these data sets are mapped in Electronic Supplementary Information (ESI) Spatial Distribution of Data Set.

2.1. Wadeable Streams Assessment

We selected WSA as the source of water quality indicators over EPA's separate assessments of lakes [23], coastal waters [24] and wetlands [25] because most people live closer to a stream than to a lake, coast or wetland. Availability was also an issue: by early 2013, lake assessment data collected in 2009, coastal data collected in 2010, wetland data collected in 2011, and rivers and streams data collected in 2008–2009 had not been released [26]. None of the data sets used in our analysis details the specific water bodies experienced by particular individuals, but these spatial relationships are essential to the analysis. Following [27], we assume the quality of nearby water bodies is salient for most people and limit the analysis to locations where both WSA and GSS or WSA and NSRE observations are present [28].

The WSA collects more than 100 indicators of water chemistry, physical conditions, and benthic invertebrate communities at specific locations on specific dates. There is considerable multi-collinearity between water quality indicators. WSA data do not include composite ratings for stream segments so, for purposes of this study, we selected three subgroups of often-used indicator variables that are not intrinsically correlated:

$WSA^{chemical} = [\text{Phos}, \text{NO}_3, \text{SDO}]$

$WSA^{habitat} = [\text{RHA}, \text{Turb}]$

$WSA^{biotic} = [\text{EPT}]$

Variable definitions appear in Table 1. The WSA variants reflect different dimensions of water quality [29]. The chemical measures (NO₃, Phos and SDO) are associated with nutrient and organic pollution but require laboratory analyses. The physical indicators (RHA and Turb) are visually-apparent. The biotic index (EPT) measures abundance of three groups of aquatic macroinvertebrates that are sensitive to a variety of habitat and water quality modifications and whose presence indicates high habitat quality. Like the chemical indicators, EPT requires specialized expertise and is not easily discerned. Within and between the three categories, the indicators are correlated but not collinear. There is debate about which indicator is most meaningful to people [30].

As noted above, WSA data are representative at ecoregional scale. Water quality, as well as its dominant stressors, varies locally and between ecoregions. Preferences may also differ regionally as people select locations based in part on environmental quality [31]. Including ecoregions as variables enables testing for regional differences in relationships between attitudes, behavior and water quality. Finding such differences would indicate that generalizable insights require observation and analysis at large scale.

2.2. General Social Survey

GSS biennially collects detailed demographic and activity data from households through in-person interviews using a randomized full-probability design representing the nation's demographic profile [32] (ESI Spatial Distribution of Data Sets). Once each decade, GSS adds questions about environmental attitudes and behaviors [19]. The only questions specifically pertinent to water appear in the behavioral and attitudinal variable definitions in Table 1. The hypothesized effects of each variable are indicated.

2.3. National Survey on Recreation and the Environment

The NSRE uses random-digit-dial telephone interviews administered on a rolling basis over multiple years—most recently, 2005–2009—to collect demographic and attitudinal information and data on boating, fishing, swimming, and waterside activities [33]. The sample is statistically representative of the US adult population. To protect privacy, the data are aggregated to polygon scale (minimum: zip code). NSRE collects information on type of water activity but not location. Aggregation of individual NSRE responses further obscures locational relationships. Table 2 lists the NSRE variables used here and their hypothesized effects on activities.

Table 1. Variables used in Case Study 1.

Variables ^a		Description and [Hypothesized effect on dependent variables ^b]	Source
Behavioral	h2oless	“How often do you choose to save or re-use water for environmental reasons? (1) Never, (2) Sometimes, (3) Often, (4) Always.”	[19]
Attitudinal	prcvdanger	“In general, do you think that pollution of America’s rivers, lakes, and streams is (1) not dangerous at all for the environment, (2) not very dangerous, (3) somewhat dangerous, (4) very dangerous, (5) extremely dangerous for environment?” [+]	[19]
	grntaxes	“How willing would you be to pay much higher taxes in order to protect the environment? (1) Not at all willing, (2) Not very willing, (3) Neither willing nor unwilling, (4) Fairly willing, (5) Very willing” [+]	[19]
	age	Age of respondent [+/-]	[19]
Demographic	children	“How many children have you ever had? Please count all that were born alive at any time (including any you had from a previous marriage).” [+/-]	[19]
	race	What race do you consider yourself? (0) Non-Caucasian [†] , (1) Caucasian [+/-]	[19]
	sex	Dummy variable: (0) Female [†] , (1) Male [+/-]	[19]
	income	Middle values of 25 annual family income intervals specified in GSS variable INCOME06; highest income category is open-ended and represented here by 110% of its lower-bound. [+/- for prcvdanger and h2oless; + for grntaxes]	[19]
	incomesq	(income) ²	
	polviews	“We hear a lot of talk these days about liberals and conservatives. I’m going to show you a seven-point scale on which the political views... Where would you place yourself on this scale? (1) Extremely liberal [†] , (2) Liberal, (3) Slightly liberal, (4) Moderate, middle of the road, (5) Slightly conservative, (6) Conservative, (7) Extremely conservative” [+/-]	[19]
	grngroup	“Are you a member of any group whose main aim is to preserve or protect the environment? (0) No [†] , (1) Yes” [+]	[19]
	popdens	Population per square mile [+/-]	[34]
Water Quality Indicators	RHA	Rapid Habitat Assessment Score Mean [- for prcvdanger and grntaxes; +/- for h2oless]	[18]
	Turb	Turbidity, measured by Nephelometric Turbidity Units (NTU) [- for prcvdanger and grntaxes; +/- for h2oless]	[18]
	Phos	Phosphorus (µg/l) [+]	[18]
	NO3	Nitrate (µeq/l) [+]	[18]
	SDO	Stream Dissolved Oxygen (mg/L) [-]	[18]
	EPT	% of individual benthic macroinvertebrates comprised of <i>Ephemeroptera</i> , <i>Plecoptera</i> and <i>Trichoptera</i> [- for prcvdanger and grntaxes; +/- for h2oless]	[18]
Regional	region	Ecoregion dummy variables: region1-Northern Appalachians, region2-Southern Appalachians, region3-Coastal Plains [†] , region4-Upper Midwest, region5-Temperate Plains, region6-Southern Plains, region7-Northern Plains, region8-Western Mountains and region9-Xeric [+/-]	[18]

Notes: ^a Bold font indicates vector; ^b +/- indicates a two-tailed hypothesis. Indicated hypotheses apply in all three models (explaining prcvdanger, grntaxes and h2oless) unless indicated separately for different dependent variables; [†] Omitted category.

Table 2. Variables used in Case Study 2.

Variable	Description and [Hypothesized Effect ^a]	Source	
Recreation ^b	boatd	# days canoeing and boating [20]	
	waterd	# days of waterside activities [20]	
	swimd	# days swimming in lakes, rivers, ponds [20]	
	fishd	# days freshwater fishing [20]	
	dist	Distance (miles) from the centroid of NSRE zip code to the nearest WSA stream site in same zip code [-] [18,20]	
Water Quality	See Table 1 (RHA[+], Turb[+], NO3[-] and Phos [-], SDO[+], EPT[+]) [18]		
Attitudes	improtect “It is important to conserve and protect National Forest Grasslands that support water resources such as streams, lakes and watershed areas: (1) strongly disagree [†] , (2) disagree, (3) neutral, (4) agree, (5) strongly agree” [+]	[20]	
Demography	age	Age [+/-] [20]	
	sex	Sex: (0) Female [†] , (1) Male [+/-] [20]	
	race	Dummy variable: (0) Non-Caucasian [†] , (1) Caucasian [+/-] [20]	
	hhsiz	Household size [+/-] [20]	
	income	Middle values of 11 annual income intervals ; highest income category is open-ended and represented here by 110% of its lower-bound [+]	[20]
	incomesq	(income) ² [-] [20]	
Regional	See Table 1 [18]		

Notes: ^a See footnote a in Table 1. Hypothesized effects relate to both recreation participation and frequency;

^b As defined, the activity variables correspond to the “#days” variables used in Equation (4). The “Participation” variable in Equation (3) equals zero if a corresponding #days variable is zero, and 1 if the corresponding #days variable is positive; [†] Omitted category.

2.4. Data Matching

For comparability with GSS and NSRE data, WSA point data must be reduced to polygon scale. Following the protocols applied to GSS and NSRE data, we average WSA observations within a polygon and do not weight the resulting averages by the number of represented observations; GSS and NSRE data lack the information needed for weighting. Averaging may reduce variance and associated statistical power while also introducing bias [35] (*cf.* [36]).

Overall, this aggregation methodology produces a sample that is almost certainly unrepresentative of water quality and population at any spatial scale. Representativeness is further threatened by a lack of spatial congruence between data sets, as summarized in Table 3. For GSS, the smallest available unit of observation is census tract. At tract scale, GSS and WSA data overlap in four instances. At county scale—there are 129 overlaps. Only county scale provides enough matched observations for statistical analysis. NSRE and WSA data overlap in 888 zip codes and 849 counties. NSRE data are not geo-referenced to census tracts. Zip code scale provides sufficient matched observations for statistical analysis.

Table 3. Spatial congruence of observations for WSA, GSS, NSRE.

Data source	Number of observations	Zip codes containing observations	Census tracts containing observations	Counties containing observations
WSA	2,035	1,339	1,197	893
GSS	2,044 ^a	-- ^a	366	196
NSRE	103,770 ^a	19,959 ^b	-- ^a	2,931 ^b
WSA-GSS Overlap	-- ^a	-- ^a	4	129
WSA-NSRE Overlap	-- ^a	888	-- ^a	849

Notes: ^a Data not available at this scale to protect respondent privacy or because the data collector does not report at the noted spatial scale; ^b 4033 NSRE observations that do not include a zip code are excluded from the zip code count but included in the county count.

To test whether the geospatially-matched observations faithfully represent the raw samples, we apply Welch's T-test for equivalence of means of samples that differ in size and variance. Even using a weak critical value of $p < 0.1$, the results support equivalence for only two of 42 CS1 variables (ecoregions 8 and 9) and five of 31 CS2 variables (waterd, income, incomesq, sex, improtect-strongly agree) (ESI Methods and Results). These findings strongly suggest that the matched samples are unrepresentative of the full samples.

3. Case Study 1: Water Quality, Attitudes, and Conservation Behavior

3.1. Model

To reiterate, Case Study 1 concerns relationships between measured water quality, environmental attitudes, and water conservation efforts. We use WSA and GSS data to explore these questions at national scale. Although 129 observations result from the matching process, 32 of those observations lack one or more data values. The remaining 97 complete observations are used in the analysis. The first step tests for spatial autocorrelation in the data. Very low and insignificant values of Moran's I statistic [37] fail to confirm spatial autocorrelation. This outcome, particularly for GSS data, almost certainly reflects a lack of spatial coverage.

To examine these questions, we adopt a simple linear model (e.g., [38]) augmented with a quadratic income term to allow for diminishing marginal utility of money—a widely-observed economic phenomenon (e.g., [39,40]). Employing variables defined in Table 1, the first model addresses RQ2:

$$\text{h2oless} = \beta_0 + \beta_1 \text{attitude} + \beta_2 \text{age} + \beta_3 \text{children} + \beta_4 \text{race} + \beta_5 \text{sex} + \beta_6 \text{income} + \beta_7 \text{incomesq} + \beta_8 \text{polviews} + \beta_9 \text{grngroup} + \beta_{10} \text{popden} + \beta_{11} \text{region} + \beta_{12} \text{WSA} + \varepsilon_1 \quad (1)$$

For polviews and region, each coded as multiple dummy variables, one category in each case must be excluded to prevent linear dependence. We omitted extremely liberal for polviews and region 3. The excluded variables are embedded in the constant and become the benchmarks for comparison to the other regional and political categories. Attitude includes two categorical indicators from GSS: prcvdnger and grntaxes. These variables are coded as levels using the integers listed in their definitions.

Many studies find codetermination of environmental attitudes and environmental behaviors (e.g., [41]), suggesting the need to test for endogeneity. Indeed, Wu-Hausman tests indicate

endogeneity between each of the water quality indicators and both attitude vectors (F statistics were 0.00323, 0.00217 and 0.00176 for regressions with chemical, physical and biotic measures respectively). We therefore implement a two-stage instrumental variable estimation strategy [38]. The first stage explains attitude formation and the second stage explains behavior as postulated in RQ2. To identify an instrumental variable, we must find a variable that affects attitude but not behavior directly. Loosely following Ki and Hon [42], we assume that political views affect attitudes toward water quality data and the attitudes in turn influence behavior. Equation (1) (modified as explained below) comprises the second stage while the following equation is estimated in the first stage:

$$\text{attitude} = \alpha_0 + \alpha_1\text{age} + \alpha_2\text{children} + \alpha_3\text{race} + \alpha_4\text{sex} + \alpha_5\text{income} + \alpha_6\text{incomesq} + \alpha_7\text{polviews} + \alpha_8\text{grngroup} + \alpha_9\text{popden} + \alpha_{10}\text{region} + \alpha_{11}\text{WSA} + \varepsilon_2 \quad (2)$$

The predicted values from Equation (2) substitute for the raw attitude variables in Equation (1). The predicted values of the attitude variables are continuous, so a continuous variable replaces each integer-valued level in stage 2. This substitution may be a source of statistical error [43]. Most econometric software packages fail to provide means of correction. In comparison to data limitations, we believe this statistical issue is of minor importance. The revised Equation (1) addresses RQ1. Since Equations (1) and (2) include the same independent variables, apart from attitude, for identification of Equation (1), one variable that is correlated with attitude but not with other behavioral variables must also be excluded. We exclude polviews. Using Stata statistical software, each of the resulting models is estimated separately for each of the three versions of WSA. We use ordered probit estimation for both models [44].

3.2. Results

Table 4 summarizes the results in a form designed to highlight robustness across the different water quality models. (Complete detailed results appear in ESI Methods and Results). Robustness is the focus because demographic influences that are discriminately measured by the data should persist across models irrespective of the particular water quality measure used. F-tests support the inclusion of ecoregional dummies in all equations, although they are rarely individually significant. (F-statistics generated with and without the ecoregion dummies for regressions with chemical, physical and EPT measures and were 221.24, 345.3 and 147.6, respectively. These are all greater than the 5% critical value.) The stage one model for grntaxes failed to converge for any of the water quality vectors so there are no results to report and grntaxes is excluded from the h20less model. For the other two models, variables for which the null hypothesis is never rejected at the 0.05 significance level or better do not appear in Table 4.

Except for the ecoregional variables in the prcvdangerequation, the few independent variables that are significant in the two equations are consistently significant across the different water quality models. However, the results are not very informative about the motivating research questions. For prcvdangerequation, none of the water quality metrics is significant and evidence of specific regional differences is meager. For h20less, no water quality indicators are significant and evidence of specific regional differences is modest. The weakness of these results probably reflects the information losses associated with county-level aggregation, the small number of matched and complete observations and the limiting assumption, made in the absence of explicit information in the survey about respondents'

exposure to water resources, that co-located stream quality is dispositive. In addition, h2oless reflects quantity choices for treated household water while the quality measures are for *in-situ* water that may or may not be associated with household supply. The relationship between the two is tenuous. However, h2oless is the only water-relevant behavioral variable available in GSS.

Table 4. CS1: WSA models for which variable is significant at $p \leq 0.05^a$.

Variable ^b	First stage: Attitude ^c	Second stage: Behavior
	prevdanger	H2oless
age	-	A ⁺
sex	-	A ⁻
income	A ⁺	-
incomesq	A ⁻	-
grngroup	A ⁻	A ⁺
polviews-liberal ^d	A ⁻	na
Region ^e	C, P	A ²
R ^{2f}	0.49	0.49
N	97	97

Notes: ^a Key: B = biotic, C = chemical, P = physical, A = all (B,C, and P) measures taken from WSA; + indicates positive effect, - indicates negative effect; ^b Variables never significant at $p \leq 0.05$ are not shown. Constants are significant in all models; ^c Stata econometric software fails to converge on a solution for any models of grntax; ^d Omitted from the second stage to enable identification. Polviews-conservative is significant and negative at $p \leq 0.10$ in all prevdanger models; ^e Indicates at least one of the ecoregional dummy variables is independently significant at $p \leq 0.05$ in the indicated model(s); if more than one, the number is indicated by the superscript. Signs are not important; ^f The first and second stages are solved simultaneously for each water quality vector so the R² values are identical for the equations estimated with a particular vector of WSA measures but differ across water quality vectors.

4. Case Study 2: Water Quality and Recreation Activities

4.1. Model

Case Study 2 (CS2) explores water quality effects on recreation participation? Answers would help in the design of information programs for recreators. The demand for access to streams for recreation has increased dramatically over the past 50 years [45]. In 2004–2005, only 53% of US stream miles were in good or fair condition while 42% were in poor condition [18]. Numerous studies have explored effects of environmental quality on recreation behavior at small scale (e.g., [46]). To explore it at large scale, testing for regional differences, we use data from the 2009 round of the NSRE [20], the 2010 Census [34] and the 2004–2005 WSA [18] (Table 2). Since many water-based recreational activities occur in streams, rivers and lakes, *in-situ* water quality seems more likely to be influential than in the case of household water conservation.

NSRE does not provide information about the distance or cost of reaching and using recreational sites. The recreation demand literature [47] frequently assumes access costs are proportional to distance between residence and recreational site. However, the NSRE data do not reveal distance. We have only the specific locations of WSA observations and the centroids of resident zip codes. The dist variable measures the linear distance between the centroids and the closest WSA observation site. Its significance would imply that proximity to a source of water quality information influences recreation

behavior. Variable *improtect* is the NSRE attitudinal variable most closely associated with water. *Improtect* is coded as levels using the integers listed in its definition. The fact that it is tangential to the question of interest is further illustration of the difficulty of adapting data collected for other purposes to study human interactions with water.

We apply a two-step hurdle model common in the recreation demand literature [48]. The frequency of participation is conditional on the initial binary decision to participate. Participation is estimated with a logistic choice model, shown in Equation (3), while frequency is estimated with the count model in Equation (4):

$$\text{Participation} = \gamma_0 + \gamma_1 \text{age} + \gamma_2 \text{sex} + \gamma_3 \text{race} + \gamma_4 \text{hhsz} + \gamma_5 \text{income} + \gamma_6 \text{incomesq} + \gamma_7 \text{impprotect} + \gamma_8 \text{dist} + \gamma_9 \text{WSA} + \gamma_{10} \text{region} + \gamma_{11} \text{region} * \text{income} + \gamma_{12} \text{region} * \text{distance} + \varepsilon_3 \quad (3)$$

$$\# \text{days} = \delta_0 + \delta_1 \text{age} + \delta_2 \text{sex} + \delta_3 \text{race} + \delta_4 \text{hhsz} + \delta_5 \text{income} + \delta_6 \text{incomesq} + \delta_7 \text{impprotect} + \delta_8 \text{dist} + \delta_9 \text{WSA} + \delta_{10} \text{region} + \delta_{11} \text{region} * \text{income} + \delta_{12} \text{region} * \text{distance} + \varepsilon_4 \quad (4)$$

The count model assumes a negative binomial distribution to correct for possible over-dispersion [37].

We consider fishing, swimming, boating and viewing activities separately due to correlations between them. Regional effects are included both directly and in interaction with income and distance. For each activity, we separately test the three versions of WSA.

4.2. Results

Summary results appear in Table 5 (details in ESI Methods and Results). Wu-Hausman tests for endogeneity of *improtect* and the recreation variables reject that possibility in all cases. Due to nonparticipants, sample sizes for frequency decisions are smaller than for participation. The frequency models are particularly weak, with significant independent variables found only in the case of swimming.

Compared to CS1, the larger data set for CS2 allows more extensive investigation of regional effects, including interactions with demographic variables. F-tests support inclusion of regional dummies, both directly and interacting with other variables. (F-test statistics were generated with and without the ecoregional dummies for regressions with Chemical, RHA and EPT measures and were valued at 211.56, 223.19 and 198.45, respectively, supporting rejection of the null hypothesis that adding the ecoregional dummies adds no significant information.) As in CS1, only a few regional terms (both direct and interactive) are individually significant, but most of the significant variables are robust across water quality models. Among WSA indicators, the single finding of significance at the 0.05 level or better is EPT for swimming frequency. Thus, concerning RQ3, the results imply that measured water quality rarely influences water recreation decisions and that evidence of regional differences in those decisions exists, more so for swimming frequency than for the other behaviors.

Where demographic variables are significant, they are robust across different measures of water quality. The limited influence of income and its negative effect where significant are contrary to expectation.

Once again, aggregation to zip code scale sacrifices variation in these data. Perhaps more critical to the analysis of recreation behavior is the assumption that stream quality near the recreator's home is the most pertinent indicator of water conditions. The absence of information about recreational destination forces an arbitrary assumption.

Table 5. CS2: WSA models for which variable is significant at $p \leq 0.05^a$.

Variables ^c	Recreational activity ^b							
	fishd		swimd		boatd		waterd	
	Part	#days	Part	#days	Part	#days	Part	#days
age	A ⁺	-	A ⁺	-	A ⁺	-	-	-
sex	A ⁻	-	-	-	A ⁻	-	-	-
race	-	-	-	-	A ⁻	-	--	-
income	-	-	-	A ⁻	-	-	A ⁺	-
Improtect ^{d,e}	A	-	A	-	A,C	-	A	-
dist	-	-	-	-	-	-	B ⁻ ,P ⁻	-
EPT ^f	-	-	-	B ⁺	-	-	-	-
Region ^d	-	-	-	A ³ ,C,P	A	-	-	-
region*inc ^d	-	-	-	A ³ ,B,C	-	-	-	-
region*dist ^d	A	-	-	-	-	-	A ²	-
N	888	839	864	371	888	864	888	801
Pseudo-R ²	0.14	0.005–0.006	0.14	0.0001–0.005	0.14	0.005–0.006	0.14	0.005–0.006
Log-likelihood	-571	-1052	-530	-1326	-551	-1135	-244	-1033

Notes: ^a Key: B = biotic, C = chemical, H = habitat, A = all measures (B, C and H) measures taken from WSA; + indicates positive effect, - indicates negative effect, ± varies between models; ^b Part = Participation (= 0 if #days = 0; 1 otherwise); ^c Variables never significant at $p \leq 0.05$ are not shown. Constants are significant except for all models of fishd-participation and boatd-participation, and in the P model of swimd-participation; ^d At least one level or region of this categorical variable is significantly different from the omitted category in one or more models. Superscripts indicate the number of significant regional or interaction terms when more than one is significant. Signs are not important and aren't reported—see ESI Methods and Results; ^e In addition to the omitted category, strongly disagree, no observations are present in the categories of neutral and agree. ^f EPT only in B models; NO₃, Phos, and SDO only in C models; RHA and Turb only in P models.

5. Discussion and Conclusions

Overarching question 1 concerns the adequacy of existing large-scale data sets to support coupled human-environmental research. A first-order problem with the data sets used here is the absence of spatial coordination in sample construction. Collocation is the only means of inferring relatedness. Applying the collocation criterion eliminates half or more of the observations. The losses are greatest in areas where population is less dense and less sampled but observations of water conditions are abundant, such as the water-scarce plains and intermountain areas. The collocation criterion is particularly suspect for the recreation analysis. Recreators often travel considerable distances to fish, boat or pursue other water-based activities, but NSRE's lack of destination information precludes linking particular water quality observations to reported activities. Mechanistic models may be available to extrapolate water quality information to non-sampled locations [49,50], thereby making it applicable to more of the social data, but no such methods are available to determine where people form impressions of environmental quality or where they recreate.

Another critical weakness of these analyses is the unrepresentativeness of the matched subsamples, so inference cannot be generalized. Finally, the data are collected in different years and this is another reason to question their coherence.

OQ2 concerns the importance of analysis at large scale. The overall significance of regional variables in both case studies suggests that observation at large scale makes a difference and findings of local studies should not be extrapolated to national scale. However, only a few of the regional variables are individually significant. Ecoregions may not be the best spatial classification scheme for water [51]; the existence and causes of regional differences warrant further exploration.

These case studies illustrate that fundamental coupled research questions depend on access to natural and social scientific data collected and coordinated at large scale. They also show that study of these questions is limited by data collection programs that fail to envision and support coupled research.

A coordinated program should begin with a suite of fundamental scientific questions [10,52]. The questions would explore how human interactions with water vary over context and time. For example: How does water quality and quantity influence location decisions and institutional evolution, and how do the resulting settlement patterns and formal and informal institutional arrangements in turn affect quality and quantity? Such questions would guide variable selection and probably require observation at several levels, including individuals and households, social networks and information channels, and formal institutions. A second element of the design process should focus on the scale of geospatial and temporal variation. Ecoregions represent major climatic and geophysical factors that regulate ecosystems, but they may not capture meaningful strata of human systems. For water, Hutchinson, Hayes and Schnoor [51] offer a first effort to discern joint clusterings of human, hydrographic and ecological features in their “Human Influenced Water Environmental Classes”. The inclusion of human elements—land cover, population density and water use—is novel for such an analysis and warrants further exploration. Core observations should be collected nationally while customized observations might be used for specific regions or sectors. Concerning temporal variation, observation is needed over long time spans because “social-ecological systems are dynamic and most likely dominated by incremental change interspersed with punctuated equilibria” [53]. These sources of variation require repeated observation over time. Third, the program should implement protocols for more precise spatial identification of human data that also protect privacy [54]. New “clustering” methods for characterizing a spatial unit are needed to integrate point and aggregated (for privacy) data, and to better link social and environmental data. Fourth, GIS-based web services, comparable to or integrated in the Water Data Center administered by the Consortium of Universities for the Advancement of Hydrologic Sciences (CUAHSI) [55], incorporating existing data sets as well as findings from complementary surveys, would facilitate broad community access. Fifth, depending on the scientific objectives, it would collect social data not only about individuals, but also formal and informal social institutions such as political jurisdictions, legal regimes, advocacy organizations, and social networks [1]. Finally, the proposed activity is long-term in nature and requires leadership of young and mid-career researchers able to see it through to realization.

The hypothesis that social and natural systems are co-determined is central to sustainability science and compelling for water problems. Our analysis provides evidence that current uncoordinated data collection programs in the US are unlikely to support generalizable insights about fundamental human-nature relationships that vary at large-scale. Important advances in the science of water sustainability depend on access to more coherent data.

Acknowledgements

We thank participants in the 2011 Workshop on Coupled Observation of the Water Environment: A National Survey Program, Amy Ando, Dan Brown, Nick Brozović, Patricia Champ, Mick Couper, Tom Holmes, David Keiser, John Loomis, Aaron McCright, Randy Rosenberger and two anonymous referees for helpful advice. Data from the National Opinion Research Center's GSS Sensitive Data Files and the U.S. Forest Service's NSRE were obtained under contractual arrangements designed to protect respondent anonymity and are available only from the original sources, not from the authors. The paper is based in part on work supported by the National Science Foundation under award 1038813 WSC, US Department of Agriculture/National Institute of Food and Agriculture project ILLU-470-316, and the Johnson Foundation. Any opinions, findings, and conclusions or recommendations expressed are these of the authors and do not necessarily reflect the views of the sponsors or advisors.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Braden, J.B.; Brown, D.; Dozier, J.; Gober, P.; Maidment, D.R.; Schneider, S.L.; Schultz, P.W.; Shortle, J.S.; Swallow, S.K.; Werner, C.M. Social science in a water observing system. *Water Resour. Res.* **2009**, *45*, doi: 10.1029/2009WR008216.
2. Schnoor, J.L. Water Sustainability in a Changing World. In *The 2010 Clarke Prize Lecture*; National Water Research Institute: Fountain Valley, CA, USA, 2010. Available online: <http://www.nwri-usa.org/pdfs/2010ClarkePrizeLecture.pdf> (accessed on 27 August 2013).
3. Mather, M. *Four Scenarios for US Population Growth*; Population Reference Bureau: New York, NY, USA, 2009.
4. United Nations. *World Population to 2300*, ST/ESA/SER.A/236; Department of Economic and Social Affairs: New York, NY, USA, 2004. Available online: <http://www.un.org/esa/population/publications/longrange2/WorldPop2300final.pdf> (accessed on 27 August 2013).
5. Intergovernmental Panel on Climate Change. *Climate Change 2007: Impacts, Adaptation and Vulnerability*; Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., Eds.; Cambridge University Press: Cambridge, UK, 2007.
6. Saad, L. Water pollution Americans' top green concern. Available online: <http://www.gallup.com/poll/117079/water-pollution-americans-top-green-concern.aspx> (accessed on 8 December 2013).
7. United Nations. *Millennium Development Goals Report 2012*; United Nations: New York, NY, USA, 2012.
8. US Environmental Protection Agency. *The Clean Water and Drinking Water Infrastructure Gap Analysis*, EPA-816-R-02-020; Office of Water: Washington, DC, USA, 2002. Available online: <http://www.epa.gov/ogwdw/gapreport.pdf> (accessed on 27 August 2013).
9. Wescoat, J.L., Jr; White, G.F. *Water for Life: Water Management and Environmental Policy*; Cambridge University Press: Cambridge, UK, 2003.

10. Sivapalan, M.; Savenije, H.H.G.; Blöschl, G. Socio-hydrology: A new science of people and water. *Hydrol. Proc.* **2012**, *26*, 1270–1276.
11. Dalhuisen, J.; Florax, M.R.; de Groot, H.L.F.; Nijkamp, P. Price and income elasticities of residential water demand: A meta-analysis. *Land Econ.* **2003**, *79*, 292–308.
12. Nelson, J.P.; Kennedy, P. The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environ. Resour. Econ.* **2009**, *42*, 345–377.
13. Janssen, M.A.; Anderies, J.M.; Ostrom, E. Robustness of social-ecological systems to spatial and temporal variability. *Soc. Nat. Resour.* **2007**, *20*, 307–322.
14. Hoehn, J. Methods to address selection effects in the meta regression and transfer of ecosystem values. *Ecol. Econ.* **2006**, *60*, 389–398.
15. Rosenberger, R.S.; Johnston, R.J. Selection effects in meta-analysis and benefit transfer: Avoiding unintended consequences. *Land Econ.* **2009**, *85*, 410–428.
16. Gotway, C.A.; Young, L.J. Combining incompatible spatial data. *J. Am. Stat. Assoc.* **2001**, *97*, 632–648.
17. Oregon Watershed Enhancement Board. *Issues in the Aggregation of Data to Assess Environmental Conditions*; Oregon Watershed Enhancement Board: Salem, OR, USA, 2009.
18. US Environmental Protection Agency. *Wadeable Streams Assessment*; Office of Research and Development Office of Water: Washington, DC, USA, 2006. Available online: <http://water.epa.gov/type/rs/monitoring/streamsurvey/> (accessed on 27 August 2013).
19. National Opinion Research Center. *General Social Surveys, 1972–2012: Cumulative Code Book*; University of Chicago Press: Chicago, IL, USA, 2013. Available online: <http://www3.norc.org/GSS+Website/Documentation/> (accessed on 27 August 2013).
20. National Survey on Recreation and the Environment (NSRE). *US Forest Service*; South Research Station: Athens, GA, USA, 1994–2012. Available online: <http://coastalsocioeconomics.noaa.gov/core/nsre/welcome.html> (accessed on 27 August 2013).
21. Omernik, J.M. Ecoregions of the coterminous United States. *Ann. Assoc. Am. Geogr.* **1987**, *77*, 118–125.
22. Herlihy, A.T.; Larsen, D.P.; Paulsen, S.G.; Urquhart, N.S.; Rosenbaum, B.J. Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP mid-Atlantic pilot study. *Environ. Monit. Assess.* **2000**, *63*, 95–113.
23. US Environmental Protection Agency. *National Lakes Assessment 2012: A Fact Sheet for Communities*, EPA 841-F-11-007; Office of Water: Washington, DC, USA, 2011. Available online: <http://water.epa.gov/type/lakes/assessmonitor/lakessurvey/upload/NLA-2012-Fact-Sheet-for-Communities.pdf> (accessed on 27 August 2013).
24. US Environmental Protection Agency. *National Coastal Conditions Report IV*; Office of Water: Washington, DC, USA, 2012. Available online: <http://water.epa.gov/type/oceb/assessmonitor/nccr/upload/Final-NCCR-IV-Fact-Sheet-3-14-12.pdf> (accessed on 27 August 2013).
25. US Environmental Protection Agency. *National Wetland Condition Assessment: Fact Sheet*, EPA-843-F-08-001; Office of Water: Washington, DC, USA, 2008. Available online: <http://www.fws.gov/wetlands/Documents/EPA-National-Wetland-Condition-Assessment-Fact-Sheet.pdf> (accessed on 27 August 2013).

26. US Environmental Protection Agency. *National Aquatic Resource Surveys*; US Environmental Protection Agency: Washington, DC, USA, 2013. Available online: http://water.epa.gov/type/watersheds/monitoring/aquaticsurvey_index.cfm (accessed on 27 August 2013).
27. David, E.L. Public perceptions of water quality. *Water Resour. Res.* **1971**, *7*, 453–457.
28. Another approach would use a composite measure of water qualities observed within a certain radial buffer around the polygons where GSS observations are reported. However, in the absence of any guidance in the data, the size of the buffer and scheme for weighting multiple WSA observations would be no less arbitrary than our chosen approach.
29. Abbasi, T.; Abassi, S.A. *Water Quality Indices*; Elsevier: Oxford, UK, 2012.
30. Van Damme, S.; Struyf, E.; Maris, T.; Ysebaert, T.; Dehairs, F.; Tackx, M.; Heip, C.; Meire, P. Spatial and temporal patterns of water quality along estuarine gradient of the Scheldt Estuary (Belgium and The Netherlands): Results of an integrated monitoring approach. *Hydrobio* **2005**, *540*, 29–45.
31. Bayer, P.; Timmins, C. On the equilibrium properties of locational sorting models. *J. Urban Econ.* **2005**, *57*, 462–477.
32. Marsden, P.V. *Social Trends in American Life: Findings from the General Social Survey Since 1972*; Princeton University Press: Princeton, NJ, USA, 2012.
33. Schuett, M.A.; Warnick, R.B.; Lu, J. Qualitative analysis of national outdoor recreation surveys. *J. Park Recreat. Adm.* **2009**, *27*, 46–59.
34. US Census Bureau. *United States Census 2010*; US Census Bureau: Washington, DC, USA, 2010. Available online: <http://www.census.gov/2010census/> (accessed on 27 August 2013).
35. O’Sullivan, D.; Unwin, D. *Geographic Information Analysis*; John Wiley: Hoboken, NJ, USA, 2003.
36. Hellerstein, D. Welfare estimation using aggregate and individual-observation models: A comparison using Monte Carlo techniques. *Am. J. Agric. Econ.* **1995**, *77*, 620–630.
37. Anselin, L. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agric. Econ.* **2002**, *27*, 247–267.
38. Garcia-Valinas, M.A.; McIntyre, A.; Torgler, B. Volunteering, pro-environmental attitudes and norms. *J. Socio Econ.* **2012**, *41*, 455–467.
39. Diener, E.; Biswas-Diener, R. Will money increase subjective well-being? A literature review and guide to needed research. *Soc. Indic. Res.* **2002**, *57*, 119–169.
40. Easterlin, R.A. Explaining happiness. *Proc. Nat. Acad. Sci. USA* **2003**, *100*, 11176–11183.
41. Liska, A.E. Emergent issues in the attitude-behavior consistency controversy. *Am. Soc. Rev.* **1974**, *39*, 261–272.
42. Ki, E.-J.; Hon, L.C. Causal linkages among relationship quality perception, attitude, and behavior intention in a membership organization. *Int. J. Corp. Commun.* **2012**, *17*, 187–208.
43. Bollen, K. An alternative two stage least squares (2sls) estimator for latent variable equations. *Psychomet* **1996**, *61*, 109–121.
44. Long, J.S.; Freese, J. *Regression Models for Categorical and Limited Dependent Variables Using Stata, 2nd ed.*; Stata Press: College Station, TX, USA, 2006.
45. Brown T.C. *The Marginal Economic Value of Streamflow from National Forests*; USDA Forest Service Rocky Mountain Research Station: Fort Collins, CO, USA, 2004.

46. Kaczynski, A.; Henderson, K. Environmental correlates of physical activity: A review of evidence about parks and recreation. *Leis. Sci.* **2007**, *29*, 315–354.
47. Parsons, G.R. *A Primer on Nonmarket Valuation*; Champ, P.A., Boyle K.J., Brown T.C., Eds.; Kluwer Academic Publishers: Dordrecht, UK, 2003; pp. 269–330.
48. Shonkwiler, J.S.; Shaw, D. Hurdle count-data models in recreation demand analysis. *J. Agric. Resour. Econ.* **1996**, *21*, 210–219.
49. Alexander, R.B.; Smith R.A. *Moving from Monitoring to Prediction: The Quality of the Nation's Streams*, Briefing Sheet; U.S. Geological Survey: Washington, DC, USA, 2005. Available online: http://water.usgs.gov/nawqa/briefing_sheet2.pdf (accessed on 12 December 2013).
50. Keiser, D. Yale School of Forestry and Environmental Studies, Personal Communication; 7 June 2013.
51. Hutchinson, K.J.; Haynes, D.A.; Schnoor, J.L. Human-impacted water resources: Domain stratification and mapping to determine hydrologically similar units. *Environ. Sci. Technol.* **2010**, *44*, 7890–7896.
52. Dozier, J.; Braden, J.B.; Hooper, R.P.; Minsker, B.S.; Schnoor, J.L.; Bales, R.C.; Conklin, M.H.; Derry, L.A.; Harmon, T.; Michalak, A.; *et al.* *Living in the Water Environment: The WATERS Network Science Plan*; National Science Foundation: Arlington, VA, USA, 2009.
53. Anonymous referee, quoted comment received on 16 November 2013.
54. Krumm, J. A survey of computational location privacy. *Pers. Ubiquitous Comput.* **2009**, *13*, 391–399.
55. Consortium of Universities for the Advancement of Hydrologic Sciences. Available online: <http://wdc.cuahsi.org/> (accessed on 16 November 2013).

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).