AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR021527

Using Survey Data to Determine a Numeric Criterion for Nutrient Pollution

Paul M. Jakus[1], **Nanette Nelson**[2,3], **and Jeffrey Ostermiller**[4]

[1]Center for Society, Economy, and the Environment, Utah State University, Logan, UT, USA, [2]Wyoming Survey and Analysis Center, University of Wyoming, Laramie, WY, USA, [3]Now at Flathead Lake Biological Station, University of Montana, Polson, MT, USA, [4]Utah Division of Water Quality, Salt Lake City, UT, USA

Key Points:
- We survey Utah residents to scientifically replicate a Montana study of perceptions of nuisance algae conditions
- Desirable conditions were closely matched across states but the full sample showed poor internal consistency in evaluating conditions
- Respondents with strongly monotonic preferences were far more likely to favor a more stringent standard than that of the full sample

Correspondence to:
P. M. Jakus,
Paul.Jakus@usu.edu

Citation:
Jakus, P. M., Nelson, N., & Ostermiller, J. (2017). Using survey data to determine a numeric criterion for nutrient pollution. *Water Resources Research*, *53*, 10,188–10,200. https://doi.org/10.1002/2017WR021527

Received 17 JUL 2017
Accepted 10 NOV 2017
Accepted article online 20 NOV 2017
Published online 6 DEC 2017

Abstract We present a scientific replication of a benthic algae nuisance threshold study originally conducted in Montana, but we do so using a different sampling methodology in a different state. Respondents are asked to rate eight photographs that depict varying algae conditions. Our initial results show that Utah resident preferences for benthic algae levels are quite similar to those of Montana residents, thus replicating the Montana study. For the full Utah sample, though, Cronbach's α indicated poor internal consistency in rating the photographs, so a "monotonicity rule" was used to identify respondents providing monotonic preferences with respect to chlorophyll *a* densities. Simple graphical analyses are combined with ordered probit analysis to determine the maximum desirable density of chlorophyll *a* (Chl *a*). Our analysis indicates that Chl *a* levels in excess of 150 mg Chl a/m^2 are undesirable, but the regression model suggests that those with strictly monotonic preferences were far more likely favor a more stringent standard.

Plain Language Summary Excess nutrients in water can cause eutrophication, a potentially undesirable condition that fosters the growth of nuisance algae. Unlike many other water quality problems, excess algae are easily discerned by those who recreate at streams, rivers, and lakes. EPA has recently tasked state water quality agencies with establishing a numeric criterion for excess nutrients. Some agencies have used public opinion surveys to gauge aesthetic preferences for algae when developing a chlorophyll *a*-based criterion. This approach implicitly assumes that (1) chlorophyll *a* densities closely correspond to the quantity and type of benthic algae and (2) preferences for algae are strictly monotonic, i.e., less algae are preferred to more algae. We scientifically replicate a Montana study and find that preferences of Utah residents demonstrate a remarkable similarity to those of Montana residents. However, responses for a significant proportion of the sample were not monotonic. Hence, the water quality criterion implied by survey data may be sensitive to the assumption of monotonicity. Multiple preferred criteria may emerge, each differing according to how researchers treat the monotonicity assumption. We find that respondents whose preferences for chlorophyll *a* are strongly monotonic are more likely to favor a more stringent criterion.

1. Introduction

Excess nutrients—primarily nitrogen and phosphorus—in water bodies lead to a condition called eutrophication, which can result in undesirable conditions such as nuisance algae, habitat degradation, odors, and low dissolved oxygen (DO), all of which harm sensitive aquatic life (Barica, 1981; Dodds, 2006). Standard toxicological methods used to develop water quality criteria have limited applicability to eutrophication because nutrient-related problems can occur at densities well below those that are toxic. Moreover, the effects of a given nutrient concentration can vary from place to place due to site-specific factors such as light availability (shading, turbidity), substrate characteristics, groundwater inputs, and adaptations of native fauna to low DO environments (Knowlton & Jones, 2006). Deleterious effects of excess nutrients also vary temporally, and the most severe impacts can occur at different times of the year (Burkholder et al., 2006; Stevenson et al., 2012). Eutrophication also can occur naturally, making the distinction between human-induced nutrient problems difficult to separate from innate background causes (Knowlton & Jones, 2006).

Recreational use of nutrient impaired water bodies is affected in at least two ways. First, some harmful algal blooms are toxic, with problems ranging from simple rashes to neurological disorders associated with prolonged exposure (Falconer, 1999). In many places, algae blooms can lead to beach closures or other water contact advisories (Utah County Health Department, 2016). Second, excessive growth of algae and macrophytes from eutrophication can entangle swimmers, boat propellers, and fishing lines, as well as being aesthetically displeasing (Dodds et al., 2009; Dodds & Welch, 2000; Suplee et al., 2009). The conspicuous nature of nuisance algae also distinguishes nutrient pollution from other aquatic pollutants whose effects on water quality may not be obvious to nonscientists. The visible effect of nutrient pollution lends itself to the development of aesthetic-based recreation criteria because the criteria can be directly linked to public preferences.

The United States Clean Water Act (CWA) authorizes states to create water quality criteria that protect the characteristics of water contributing to human values such as culinary water supply, irrigation supply, fisheries, aesthetics, and recreation uses. States may determine a water body's designated uses based on naturally occurring chemical, physical, geographic, and biologic characteristics. Nearly all surface waters are assigned recreation and aquatic life designated uses in accordance with the CWA's goal of managing for the "protection and propagation of fish, shellfish, and wildlife and provid[ing] for recreation in and on the water" (CWA §101(a)(2)). Water quality criteria define the benchmarks needed to achieve all of the uses assigned to a water body. The U.S. Environmental Protection Agency (EPA) uses both numeric and narrative criteria when implementing the CWA, but EPA has recently "...recognized the importance of having numeric criteria for both phosphorous and nitrogen and has urged states and tribes to prioritize waters for development of numeric criteria" (Environmental Protection Agency, 2014, p. 7). Because of algae's sensitivity to nutrient pollution as well as their connection to aquatic life, drinking water sources, and recreational designated uses, algal biomass serves as a useful assessment endpoint for the development of numeric nutrient criteria (Paul et al., 2017).

Fortunately for state water quality managers, some research into the relationship between algal level and aesthetic impairment has already been completed. Dodds and Welch (2000) reviewed literature related to aesthetic impairments of filamentous algae in flowing waters and concluded that nuisance levels of benthic algae occur somewhere between 100 and 200 mg/m^2, as measured by benthic chlorophyll *a* (Chl *a*). An earlier study conducted by Horner et al. (1983) suggested that a benthic biomass of 150–200 mg/m^2 represented a nuisance condition. Given the subjective nature of what constitutes too much benthic algae in rivers and streams, Suplee et al. (2009) sought public opinion in Montana's development of numeric nutrient criteria for wadeable streams and rivers. The authors showed respondents eight photographs, each depicting a different benthic algal Chl *a* density (ranging from 40 to 1,280 mg/m^2), and they found that benthic algae levels in excess of 150 mg/m^2 Chl *a* were not desirable. A similar effort at obtaining public input on algal visual cover thresholds was conducted in West Virginia as part of that state's effort to develop numeric criteria (Responsive Management, 2012). Respondents were asked for their reactions to photographic images of seven levels of algae coverage ranging from 4% to 65%. Nearly half of respondents indicated an algal coverage of more than 25% was unacceptable. Researchers in New York State also opted for an aesthetic-based approach in assessing the impact of excess nitrogen and phosphorus on recreational use in rivers and streams (Smith et al., 2015). Here though, the perceptions of water quality monitoring field crews were used to evaluate whether or not the water body supported designated recreational uses. The authors found that recreational use ratings provided by field crews consistently matched water chemical variables and measures of biological condition.

After being charged with developing numeric criteria for nutrient impairment of the state's water bodies, the Utah Division of Water Quality (UDWQ) commissioned a survey estimating public support for and willingness to pay for a nutrient reduction program. UDWQ also sought a nuisance benthic algae standard that could be based, in part, on preferences of the general public; in doing so, UDWQ was interested in replicating, as closely as possible, the survey approach used by Suplee et al. (2009) in Montana (hereafter, referred to simply as "Suplee").

This study uses the same photographs as those selected by Suplee but with a different sampling methodology in a different state; by definition, our study is an example of scientific replication (Camfield & Palmer-Jones, 2013; Hamermesh, 2007). The role of replication in the social sciences has received more attention in recent years yet, with the notable exception of psychology and experimental economics, social science researchers have not been especially active in assessing study replicability (Camerer et al., 2016; Camfield & Palmer-Jones, 2013; Freese, 2007; Hamermesh, 2007). In addition, our interest in replicating the Suplee study is that it will help water quality managers assess the validity of the photographs in conveying the key components of undesirable algae levels to the public. Despite the scientific studies describing benthic algae levels in streams and rivers that may impair designated uses, Suplee argues "some type of assessment of the public's opinion on the matter is clearly warranted." Confirming Suplee's methodology through scientific replication will lend credibility to including public opinion when making water resource management decisions, especially if surveys were to reveal a consensus as to what constitutes a desirable level of benthic algae in similar ecoregions. We then extend the analysis to examine the effects of monotonicity of respondent preferences and estimate a statistical model linking desirable Chl *a* levels to respondents' water quality concerns, recreation use, and demographic characteristics.

2. Methods

2.1. Survey Design and Preference Elicitation

The survey design and data collection effort is described in detail elsewhere (Jakus et al., 2013; Nelson et al., 2015), but we provide a brief overview here. The design of the survey instrument was based on three focus groups conducted in three Utah cities. Given the target population was every household in Utah, our sampling frame was the U.S. Postal Service Delivery System File. Data collection followed a modified Dillman Tailored Design Method (Dillman, 2000) with up to four contacts—an advance letter followed by up to three mailings of the survey itself—during the Summer and Fall of 2011. After adjusting for nondeliverable surveys, the raw response rate was 25.3% ($n = 625$ completed surveys).

The advance letter and the cover letter for the additional mailings were printed on Utah Division of Water Quality letterhead and signed by the agency director. The text of each letter encouraged respondents to consider their answers carefully because the results would be used ". . .in making balanced decisions about how Utah's lakes and rivers are managed." This letter follows modern survey practice in constructing survey materials that reinforce the *policy consequentiality* of the survey. Numerous studies have shown that the answers of people who believe a survey is consequential can differ from those who think the survey will have little or no effect on policy (Carson & Groves, 2007; Herriges et al., 2010; Lewis et al., 2016; Vossler et al., 2012). Surveys that establish policy consequentiality goes beyond persuading recipients to take the survey seriously. Theoretic modeling has shown that differences in responses arise because those who believe a survey is consequential are more likely to adopt a dominant strategy of truthfulness when answering questions.

Nonresponse bias was evaluated using propensity score adjustment (Groves, 2006; Rosenbaum & Rubin, 1983); the propensity score variable was insignificant in our modeling, indicating that nonresponse bias was negligible (Jakus et al., 2013, Appendix D). In other words, the households who chose to respond to our survey did not vary systematically from households choosing not to respond. A comparison of demographic data from the 2010 Census on Utah households indicated differences in household size, race, and income. Sampling weights were constructed using the raking command, *ipfweight*, in Stata Version 12. Raking calculates survey weights using a stepwise approach until known population margins are achieved. Unless otherwise noted, sampling weights are applied to all of the Utah data.

Participants were shown the same eight color photographs used by Suplee, each of which had the stream bottom visible but with varying densities of benthic algae (see Appendix A for photos). Suplee selected the photos to represent the range of benthic algae conditions a user may encounter at a river in Montana during summer peak algal growth. At each site, 10–20 replicate benthic algae samples had been taken so the Chl *a* density was known for each photograph but not shown to the participants. Chl *a* densities in the photos ranged from 40 to 1,280 mg Chl a/m^2. Survey pretests found that the order in which the photographs were presented had no effect on the distribution of responses so long as the order was random. The Montana mail survey had two photos per page, which were included as part of a five page survey pamphlet; the two choice categories (desirable/undesirable) were placed adjacent to each photo.

Although the photos in the Utah survey were presented in the order selected by Suplee, the structure of our preference elicitation departed from the Montana mail survey in the following ways: (1) respondents were presented with four photos on each side of a single-page glossy insert, (2) the survey rating question

10.1002/2017WR021527

(desirable/undesirable) was included in the separate survey booklet, and (3) the wording of the desirability question was changed. The desirability question in the Utah survey was as follows:

> Please review the photos of algae in rivers on both sides of the one-page insert included in this survey. For each photograph on the insert tell us if the level of algae would be desirable or undesirable for YOUR most common uses of rivers, if any. There are no correct answers; this is your opinion only. *Fill in one bubble for each number.*

In addition to the presentation of the preference question, the key methodological differences between the Suplee study and this study concern the sampling frame and the type of respondents contacted. While the Utah study used a single random sample aimed at characterizing a general population, Suplee evaluated two groups of people. One group was contacted by mail using addresses drawn from Montana's Centralized Voter Registration files and that could be considered akin to our general population survey Utah residents. Suplee's second group was intercepted while engaged in river-based recreation on Montana Rivers. This group of respondents is clearly not representative of a general population but, rather, includes specialists who may be more sensitive to differences in the density of Chl *a*. Our Utah general population survey asked respondents about water-based recreation activities. This allows us to limit the sample to only those using rivers for recreation (defined as those making at least one recreation trip to rivers in the previous 12 months), thus permitting us to compare our river users to Suplee's on-river intercept survey.

2.2. Aesthetic Preferences and Monotonicity

Many researchers have used photographs to measure respondents' aesthetic perceptions of various resources (e.g., forests, lakes, or rivers), and they examine how these perceptions align with objective measures used by experts or planners. In general, researchers have found a fairly remarkable degree of correspondence between preferences and observable measures of environmental conditions (e.g., Junker & Buchecker, 2008; Le Lay et al., 2013; Meitner, 2004). However, using photographs to search for a single threshold implicitly assumes there is a single benthic algal density that divides preferences for benthic algae into two groups; one side of the divide reflects levels deemed desirable by a majority of the sample, while the other side shows levels deemed undesirable. This approach implicitly assumes that (1) Chl *a* densities closely correspond to the quantity and type of benthic algae and (2) preferences for benthic algae are strictly monotonic, i.e., less algae is preferred to more algae. If these conditions hold, ordering of photographs from the lowest Chl *a* density to the highest should yield monotonically decreasing percentages of the sample declaring the photos as desirable.

For reasons discussed below, preferences for algae conditions may not be monotonic in Chl *a* density, so we must also consider the strength of monotonicity in responses. Table 1 provides seven examples from our data set. We define "strongly monotonic" responses as those that are strictly monotonic in Chl *a*. For example, respondent #1004 identified the photograph with a density of 150 mg/m^2 as desirable and the photograph at a density of 200 mg/m^2 as undesirable, with all photographs with less than 150 mg/m^2 rated desirable and all greater than 200 mg/m^2 rated undesirable. Similarly, respondent #1040 shows a similar

Table 1
Monotonicity and Coding Examples

	Chl *a* density (mg Chl *a*/m^2)									
ID	40	110	150	200	240	300	400	1,280	Highest desirable Chl *a*	Monotonic?
1004	D	D	D	U	U	U	U	U	150	Strong
1040	D	D	D	D	D	D	U	U	300	Strong
1374	U	U	U	D	D	D	D	D	1,280	Strong
1067	D	D	D	U	D	U	U	U	240	Weak
1123	D	D	D	U	D	D	U	U	300	Weak
1001	D	D	U	U	U	D	U	U	300	Not
1060	D	D	U	D	U	D	U	D	1,280	Not

Note. D, desirable; U, undesirable.

Table 2
Highest Desirable Chl a Level, Utah (n = 555)

Level (mg Chl *a*/m^2)	Ordered probit dependent variable value	*N*	Strongly or weakly monotonic (%)	Strongly monotonic (%)
<40	9	7	100.0	100.0
40	8	13	100.0	100.0
110	7	129	100.0	99.2
150	6	139	100.0	96.4
200	5	22	90.9	72.7
240	4	112	78.6	16.1
300	3	41	39.0	0.0
400	2	47	14.9	2.1
1,280	1	45	48.9	31.1

pattern but with the threshold at 300 mg/m^2. Our sample also includes two respondents who demonstrate a strongly monotonic preference for increasing levels of algae (e.g., respondent #1374).

We also recognize the perceptual difficulty of the rating task and allow for minor departures from monotonicity. For example, respondent #1067 rated 200 mg/m^2 as undesirable and 240 mg/m^2 as desirable, with everything below 200 mg/m^2 desirable and everything above 240 mg/m^2 as undesirable. A single departure could occur at any point (as with respondent #1123 at 200 mg/m^2), but if only one departure from monotonicity occurs, relative to the highest desirable density, a respondent's preferences are classified as "weakly monotonic." Finally, some respondents may make multiple departures from monotonicity in preferences for Chl *a*. Respondent #1001 rates the two lowest densities as desirable, then the next three levels as undesirable (150, 200, and 240 mg/m^2), followed by another desirable outcome at 300 mg/m^2. Given that two or more departures from monotonicity have occurred relative to the highest desirable density, the respondent is classified as "not monotonic." Multiple departures from monotonicity can be problematic if researchers are to use consumer preferences to help determine a nutrient criterion; for example, the preferences for respondent #1060 appear to be ill formed, at best, and completely random, at worst. In our assessment of strength of preference ordering, 60.8% of the weighted sample are strongly monotonic, 18.8% are weakly monotonic, and 20.4% were not considered to be monotonic. Thus, 79.6% of the sample was classified as either strongly or weakly monotonic. All analysis was completed with Stata 15 (StataCorp, 2017).

2.3. Regression Modeling

Assessing an individual's preferences as depicted in Table 1 allows us to identify the highest acceptable level of Chl *a* for each respondent. Respondents' highest acceptable level of Chl *a* was converted to an ordinal scale taking values between one and nine, where a highest preferred level of 1,280 mg Chl *a*/m^2 is assigned a code of 1 and highest preferred level less than 40 mg Chl *a*/m^2 is assigned a code of 9 (see the examples in Table 1 and the final distribution of ordinal codes in Table 2). The ordinal coding of the dependent variable necessitates an ordered response model. We use an ordered probit model to estimate the probability of a respondent choosing a level of Chl *a* as being desirable (1, 2, 3, etc.) based on a suite of defining characteristics: a respondent's stated concerns about water conditions, water-based recreation activities, and demographics. Table 3 summarizes the descriptive statistics for key variables used in the model.

Ordered probit models estimate the probability of any given response, and all response category probabilities must sum to 1 (Greene, 2008, pp. 831–835). The sign of any given coefficient informs the analyst of the general direction of the relationship, but not the actual effect on any given response category. The coefficients have no simple interpretation because the entire probability density shifts to the left or to the right depending on the sign estimated coefficient and the direction of the change in the variable's value. Hence, a positive coefficient for a dummy variable that changes from a value of zero to a value of one shifts the

Table 3
Descriptive Statistics

Variable	Definition	Mean	Std. dev.	*n*
Prevent future algae blooms	Preventing future algae blooms is "highly important" (1 = yes, 0 = no)	0.462	0.499	549
Prevent future WQ reductions	Preventing future reductions in water quality is "highly important" (1 = yes, 0 = no)	0.699	0.459	548
Lake user	Has visited lake at least once in past 12 months (1 = yes, 0 = no)	0.683	0.466	551
River user	Has visited river at least once in past 12 months (1 = yes, 0 = no)	0.623	0.484	543
Female	1 = female, 0 otherwise	0.423	0.494	550
College	Graduated with 4 year degree (1 = yes, 0 = no)	0.524	0.500	549
Age	Age, in years	49.378	16.833	546
Income	Income ($1000)	69.678	43.840	535
Strongly monotonic	1 = yes, 0 otherwise	0.608	0.489	555

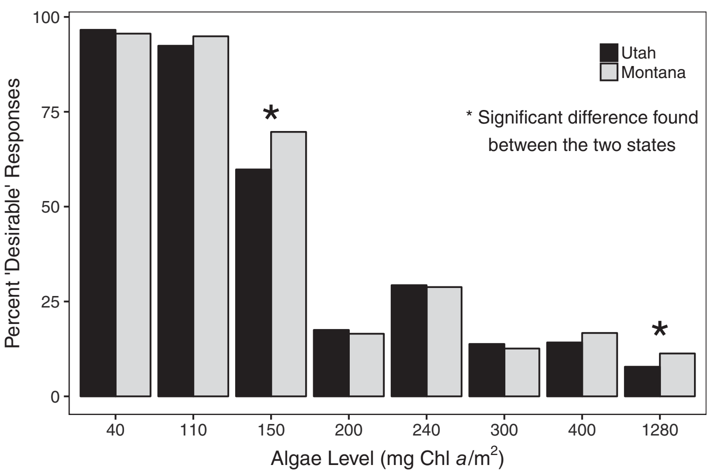

Figure 1. Desirable Chl *a* levels, Utah general population survey versus survey of Montana registered voters.

probability density to the right. That is, we can conclude that the probability of this respondent having selected response category 1 (1,280 mg/m^2) has decreased while the probability of having selected response category 9 (<40 mg/m^2) has increased. Similarly, if the model estimates a negative coefficient for the same variable, then the probability density would shift to the left (response category 1 is more likely and response category 9 is less likely). In neither case do we know what happens to the probability of being in response categories 2–8. Instead, we must conduct marginal effects analysis to determine the effect on any given response category. Just as the sum of the probabilities across all nine response categories must add up to 1, the cumulative sum of the marginal changes in response probability must add up to 0.

3. Results

3.1. The Distribution of Desirable Algae Conditions Across States

The Utah study resulted in 625 returned surveys, of which 555 respondents provided ratings for all eight photos. Given the Utah survey is a general population survey, we first compare our full sample to Suplee's general population survey, the registered voter sample (see Figure 1). A clear threshold at 150 mg Chl *a*/m^2 divides the photos into desirable and undesirable conditions for a majority of respondents. Regardless of whether the data were collected in Utah or Montana, photos depicting $\leq$150 mg Chl *a*/m^2 all had significantly more than 50% of respondents rating the photo as desirable whereas those photos depicting $\geq$200 mg Chl *a*/m^2 all had significantly more than 50% of respondents rating such photos as undesirable. Comparing the Utah and Montana samples, we observe two levels (150 and 1280 mg/m^2) at which the proportions are significantly different across states ($p = 0.002$ and $p = 0.068$, respectively), otherwise the proportion of respondents rating the Chl *a* levels as desirable are statistically identical across the two samples.

Figure 2 depicts preferences for benthic algal densities among river users for both states. Since we are not attempting to characterize a general population, the Utah sample is not weighted. The limit at which the Chl *a* level is undesirable remains the same as in the general population surveys; densities greater than 150 mg Chl *a*/m^2 are perceived as undesirable. Testing differences across the states, the proportion of respondents stating photos with a given Chl *a* level were desirable were statistically identical for all but two levels (150 and 200 mg/m^2; both with $p = 0.001$).

3.2. Assessing Preference Monotonicity for Chl *a*

Figures 1 and 2 order algal levels from least (40 mg/m^2) to most (1,280 mg/m^2) using Chl *a* as a measure of benthic algae. If Chl *a* accurately captures all the dimensions of what the public cares about concerning water recreation, and if preferences are monotonic in Chl *a*, then one should observe steadily declining percentages of survey respondents declaring the photos as desirable. However, this proves to be the case for only Montana river users (Figure 2). The proportion of people rating the 240 mg/m^2 photo as desirable is greater than the proportion rating the 200 mg/m^2 photo for both general population surveys in Utah and Montana in Figure 1 and Utah river users in Figure 2. Anomalous proportions also occur in the Utah samples between 300 and 400 mg/m^2. It does not appear that aesthetic preferences for levels of Chl *a* are monotonic.

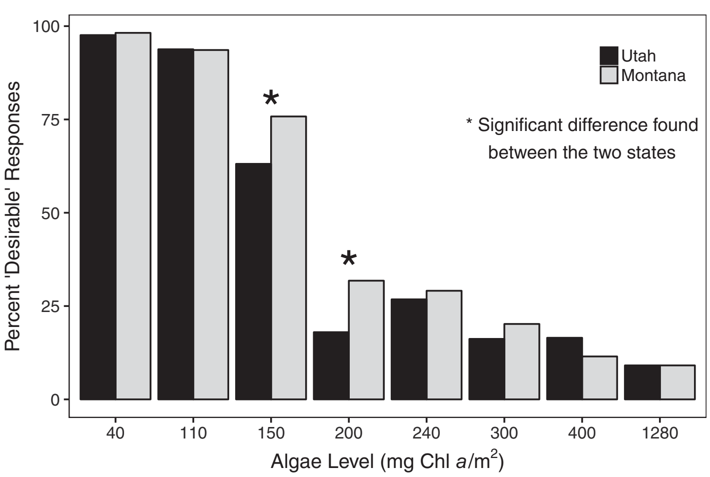

Figure 2. Desirable Chl *a* levels, Utah general population survey river users versus Montana on-river intercept survey.

Chronbach's α can be used to assess internal consistency in rating the photographs: if the eight photos all measure the same underlying construct (individual preferences for algae, calibrated to Chl *a*) one should observe a high value for the statistic. Cronbach's α for the general population Utah sample is 0.555, which denotes poor internal consistency across the photos (Table 4, column 2). In other words, some people may have preferences that are not monotonic in Chl *a* or, perhaps, factors other than Chl *a* density influenced the aesthetic

Table 4
Desirable Chl a *Levels, by Preference Monotonicity in Chl* a

Level (mg Chl *a*/m^2)	All, % desirable (*n* = 555)	Weakly or strongly monotonic, % desirable (*n* = 441)	Strongly monotonic, % desirable (*n* = 331)	All versus weakly or strongly monotonic (*p*-value)	All versus strongly monotonic (*p*-value)
40	96.6	97.1	97.3	0.628	0.514
110	92.4	93.6	93.3	0.466	0.611
150	59.8	64.0	55.0	0.185	0.172
200	17.5	16.7	14.7	0.720	0.268
240	29.3	27.3	9.6	0.504	0.001
300	13.8	7.8	4.2	0.002	0.001
400	14.2	5.8	4.2	0.001	0.001
1,280	7.8	4.7	3.9	0.046	0.016
Chronbach's α	0.555	0.667	0.718		

desirability of the scenes depicted in the photographs. Chronbach's α is 0.667 for the sample of strongly and weakly monotonic respondents, which also falls below the criterion believed to indicate acceptable internal consistency ($\alpha = 0.7$). It is only for the sample restricted to strongly monotonic respondents that Cronbach's α exceeds this cutoff value (0.718, in Table 4, column 4).

Table 4 depicts the effect of respondent monotonicity on the distribution for desirable versus undesirable Chl *a* levels. The second column shows the desirability distribution for all Utah respondents. The third column shows the distribution for all respondents classified as either weakly or strongly monotonic in Chl *a*, whereas column four shows the distribution for only strongly monotonic respondents. After adjusting for preference monotonicity, all three distributions result in the same threshold for desirability, somewhere between 150 and 200 mg Chl *a*/m^2. However, those with weakly or strongly monotonic preferences rate photos with densities equal to or greater than 300 mg Chl *a*/m^2 as significantly less desirable than the full sample (column 5). An even stronger pattern emerges in the comparison between all respondents and those that are strongly monotonic: a significantly greater proportion of strongly monotonic respondents rate all photos with a Chl *a* level equal to or in excess of 240 mg/m^2 as undesirable compared to all respondents (column 6).

3.3. Factors That Predict the Highest Desirable Level of Chl *a*

The estimated ordered probit models show that preferred Chl *a* levels are a function of stated concern about future water conditions, water-based recreation activities, and demographics (see Table 5). Our reported models highlight the effect of respondents with strongly monotonic preferences because this was

Table 5
Ordered Probit Models

Variable	Model #1 All respondents (*n* = 498)	Model #2 All respondents (*n* = 498)	Model #3 Strongly monotonic respondents (*n* = 303)
Prevent future algae blooms	0.350 (0.001)	0.405 (0.001)	0.514 (0.001)
Prevent future WQ reductions	0.368 (0.001)	0.363 (0.003)	0.330 (0.051)
Lake user	0.014 (0.914)	0.094 (0.436)	0.050 (0.750)
River user	−0.216 (0.073)	−0.292 (0.013)	−0.279 (0.085)
Female	0.300 (0.004)	0.276 (0.008)	0.442 (0.002)
College	0.013 (0.894)	−0.046 (0.641)	−0.054 (0.687)
Age	−0.003 (0.419)	0.003 (0.309)	0.006 (0.216)
Income	1.66×10^{-3} (0.154)	9.46×10^{-4} (0.409)	8.90×10^{-4} (0.556)
Strongly monotonic		2.279 (0.001)	
Log likelihood	−911.735	−742.444	−402.431
χ^2 ($\beta = 0$)	50.60 (0.001)	281.40 (0.001)	43.96 (0.001)

Note. p-Values in parentheses; based on robust standard errors. Cut values (estimated intercepts) suppressed.

the source of the largest number of statistically significant differences from the full sample. Models #1 and #2 (columns 2 and 3) include all respondents with complete data whereas those in Model #3 (column 4) use restrict the sample to only strongly monotonic respondents.

The coefficient signs for each variable were constant across all models, so our initial discussion will not distinguish between the two. Positive and statistically significant coefficients are estimated for *prevent future algae blooms*, *prevent future WQ reductions*, and *female*. *River user* was the only variable with a negative and statistically significant coefficient estimate. Model #2 (column 3) is a simple extension of Model #1 and includes a dummy variable indicating respondents those respondents whose preferences were *strongly monotonic*; the coefficient is positive and significant. All other variables (*lake user*, *college*, *age*, and *income*) were statistically insignificant in all three models.

Marginal effects analysis reveals how respondent water quality concerns, activities, or demographics affect the distribution of desirable/undesirable responses, i.e., the net effect on the probability of a given response category. Table 6 portrays quantitative estimates of the marginal effects for the three models appearing in Table 5. If an explanatory variable appearing in Table 5 does not appear in Table 6 that is because all of the marginal effects for that variable were statistically insignificant ($p > 0.10$). Similarly, an empty cell in Table 6 means the marginal effect of a given variable on that response category is insignificant.

Turning first to Model #1 and holding other factors constant, one can see that a person who believes that preventing future algae blooms is highly important (*prevent future algae blooms* = 1) is more likely to have selected response categories 6–9 (150 mg/m^2 or less) and a lower probability of having selected categories 1–5. The cumulative marginal effect for all Chl *a* densities less than or equal to 150 mg/m^2 is 13.1% more likely relative to those respondents who did not think preventing future algae blooms was highly important. Similar results hold for the other variables with positive coefficients, *prevent future WQ reductions* and *female*. The variable *river user* has a negative coefficient, and the probability a river recreationist selecting response categories 1–4 (i.e., preferring Chl *a* densities of 240 mg/m^2 or more) is cumulatively 8.0% more likely than respondents who do not recreate at rivers.

Models #2 and #3 were estimated to gauge the effect of strongly monotonic respondents on the predicted response category. Model #3 simply replicates Model #1 with a sample restricted to only those respondents classified as strongly monotonic; the results are qualitatively similar to those of Model #1 except that the

Table 6
Marginal Effects: Increase/Decrease in Response Category Probability

	Response category								
	1 (1,280)	2 (400)	3 (300)	4 (240)	5 (200)	6 (150)	7 (110)	8 (40)	9 (<40)
Model #1: All respondents									
Prevent future algae blooms	−0.051	−0.028	−0.021	−0.029	−0.002	+0.021	+0.084	+0.015	+0.011
Prevent future WQ reductions	−0.054	−0.029	−0.023	−0.030	−0.002	+0.023	+0.088	+0.016	+0.012
River user	+0.032	+0.017	+0.013	+0.018		−0.013	−0.052		
Female	−0.044	−0.024	−0.018	−0.025	−0.002	+0.018	+0.072	+0.013	+0.010
Model #2: All respondents									
Prevent future algae blooms	−0.047	−0.012	−0.008	−0.015	−0.006		+0.068	+0.015	+0.012
Prevent future WQ reductions	−0.042	−0.011	−0.007	−0.013	−0.005		+0.061	+0.014	+0.011
River user	+0.034	+0.009	+0.006	+0.011	+0.004		−0.049	−0.011	−0.009
Female	−0.032	−0.008	−0.005	−0.010	−0.004		+0.047	+0.010	+0.008
Strongly monotonic	−0.264	−0.070	−0.045	−0.084	−0.032		+0.385	+0.085	+0.066
Model # 3: Strongly monotonic respondents									
Prevent future algae blooms	−0.042			−0.037	−0.030	−0.072	+0.128	+0.031	+0.025
Prevent future WQ reductions	−0.027			−0.023	−0.019	−0.047	+0.082	+0.020	+0.016
River user				+0.020			−0.069		
Female	−0.036			−0.031	−0.026	−0.062	+0.110	+0.027	+0.021

Note. All statistically significant marginal effects shown ($p \leq 0.10$). Marginal effects for four variables (lake user, college, age, and income) were all statistically insignificant; values are suppressed for clarity. Each response category shown with its corresponding level of chlorophyll in parentheses (mg Chl *a*/m^2).

breakpoint for the marginal effects occurs at response category 7 (110 mg/m^2) instead of category 6. However, the specification that best illuminates the issue of montonicity is Model #2. The sample used for this model was identical to Model #1 (all respondents with complete data) but the specification was augmented with a dummy variable identifying *strongly monotonic* respondents. This coefficient was highly significant and it had a very large magnitude relative to other coefficients (Table 5). The large magnitude translates into a large marginal effect (Table 6).

Strongly monotonic respondents have a 38.5% greater probability of selecting response category 7 (110 mg/m^2), all else equal, relative to those who were classified as not monotonic or weakly monotonic. Cumulatively, the probability of a strongly monotonic respondent selecting category 7, 8, or 9 (preferring Chl *a* densities $\leq$110 mg/m^2) is 53.6% more likely than those respondents whose responses are not strictly monotonic.

4. Discussion

Our effort at scientific replication found a remarkable correspondence with the results reported by Suplee et al. (2009). Using the same photographs to elicit respondent preferences for benthic algal Chl *a* levels, the distribution of preferences by Utahns were very similar to those of Montanans despite clear differences in sampling frames, survey format (mail versus intercept), and even differences in the preference elicitation question. Regardless of these differences, very few statistically significant differences were found when comparing the Utah general population to registered Montana voters, or when comparing Utah river users (contacted via mail) to Montana river users (intercepted on-site). Both the Utah and Montana studies show that a desirable threshold for benthic algae (150 mg Chl *a*/m^2) for recreational use corresponds to benthic algae levels identified in the literature as "...representative of the onset of eutrophic conditions in temperate streams" (Suplee et al., 2009, p. 135). Our scientific replication in Utah of the Montana study demonstrates a notable similarity in societal preferences regarding the management of water quality for recreational use in small streams and rivers in northern temperate regions. However, one would anticipate that people's expectations and preferences are likely to vary by region because the physical environment giving rise to eutrophication will differ by region and respondents' tolerance for algae levels is likely to be conditioned by their experience within that environment.

A key question arising from our data is whether the density of benthic algae, as measured by Chl *a*, is a sufficiently good indicator of what the public cares about in regard to water-based recreation. About 20% of respondents did not exhibit monotonic preferences for Chl *a*, and a Chronbach's α calculated for the full sample confirmed a lack of internal consistency in rating the eight photographs. However, Chronbach's α for respondents who were stongly monotonic—which comprise 61% of the sample—suggest that the Suplee selection of photographs was successful in getting a majority of respondents to focus on assessing levels of Chl *a* rather than other factors.

Any of a number of factors could affect the monotonicity of responses, but here we note four possible reasons. First, preferences for benthic algae coverage may not be strictly monotonic with respect to Chl *a*. Instead, preferences may relate to other attributes of the algae depicted in the photographs, such as color and length. Filamentous algae, in particular, can be especially annoying to anglers, boaters, and swimmers given its long strands and tendency to form mats on the water surface. The algae in Photo B (240 mg Chl *a*/m^2) are brown, whereas the algae in Photo E (200 mg/m^2) are both bright green and filamentous. For some respondents, the attributes of the algae depicted in the two photos may have tipped the scale against Photo E even though it has the lower Chl *a* density.

A second possible reason one might not observe monotonic preferences in Chl *a* is that respondents' ratings could be based on compositional elements other than algae. For example, Photos A (40 mg Chl *a*/m^2) and B (240 mg Chl *a*/m^2) contain obvious nonaquatic elements (a bridge and a road cut, respectively) that may have distracted respondents away from the focus on water quality. Third, as noted by Suplee, nonexperts may have trouble discerning the distinction between Chl *a* levels, especially when the difference is relatively small. The 110, 150, 200, and 240 mg/m^2 are all separated by 50 mg/m^2 or less, thus making the exercise perceptually challenging for survey respondents. Finally, it is possible that some portion of the sample did not consider the survey to be policy consequential and therefore did not give survey questions careful consideration. A perceived lack of policy consequentiality may explain seemingly random responses such as those provided by respondent #1060 (Table 1).

Our analysis has revealed the importance of accounting for the possibility of nonmonotonicity in preferences. Though simple tabular and graphical analysis of the density level dividing the preference distribution into desirable and undesirable benthic algae levels did not change after incorporating the monotonicity rule, ordered probit analysis showed heterogeneity among respondents' tolerance for benthic algae levels. The factors influencing heterogeneity were preferences regarding future water conditions, whether one is active in river recreation, gender, and strong monotonicity. In particular, preferences of strongly monotonic respondents shift the probability density to the right by what can only be described as a huge amount. In this study, the distributional shift was not quite large enough to generate a different benthic algae threshold relative to those whose preferences were not strictly monotonic. However, the ordered probit modeling illustrates the very real possibility that other studies relying upon surveys of the general public could produce thresholds that differ depending upon how researchers choose to treat respondent preferences.

5. Conclusions

Taken as a whole, we have found a benthic algae nuisance threshold at levels in excess of 150 mg Chl *a*/m^2. In addition to being identical to the threshold identified by Suplee et al., it is also consistent with what other studies have characterized as nuisance levels (e.g., Dodds & Welch, 2000; Horner et al., 1983). The EPA recommends that states and tribes consider the use of an algal biomass indicator such as Chl *a* in developing numeric nutrient criteria. From a management perspective, the identified threshold of 150 mg Chl *a*/m^2 in the intermountain west can be assessed with field algal data collection and lab analyses routinely conducted by states. Although the public perception of algal level is not likely to be constant across the country, there may be consensus to what is desirable within similar ecoregions, as we have found for Montana and Utah. Managers in other ecoregions wishing to evaluate societal preferences to establish a nutrient criterion protective of recreational uses may identify a different limit.

Utah's DWQ is using these results in their current process of establishing a benthic algae density for aquatic aesthetics. Understanding what the public deems desirable (or acceptable) for recreational uses is critical to protecting these uses. The proposed benthic algae density, along with empirically derived Utah-specific thresholds for total nitrogen and total phosphorus, are part of UDWQ's effort to protect streams and lakes from cultural eutrophication. This suite of numeric nutrient criteria is a critical component of Utah's initial efforts to prevent degradation of aquatic life and recreation uses in headwater streams, which will be followed by the development of site-specific criteria for larger order streams across Utah.

Given that the search for a single criterion is predicated upon the assumption of monotonicity in preferences, a lack of preference monotonicity in Chl *a* could complicate the water quality manager's goal to develop a single numeric criterion. It is quite possible that an analyst could find different thresholds for different portions of the sample depending on how one chooses to implement the monotonicity assumption. Though these differences did not affect the identified threshold in this study, they do suggest that future studies examine sensitivity of policy conclusions to preferences. Finally, additional research can help clarify what survey respondents react to when evaluating photographs that depict desirable and undesirable conditions for recreation. Specifically, is the density of benthic algae driving their responses as intended by the research design, or could other factors influence their answers to survey questions?

Appendix A: Photographs Used in Suplee et al. (2009)

The photographs below were originally presented to Montana residents (Suplee et al., 2009) and then used again to gauge preferences of Utah residents (this study). The Appendix presents the photos in the exact order and layout as presented in the Utah survey, although none of thephotographs had its Chl a density identified. The order in which a photo appears is the same random order as selected by Suplee et al. (2009).

10.1002/2017WR021527

Photo A: 40 mg Chl a/m^2

Photo B: 240 mg Chl a/m^2

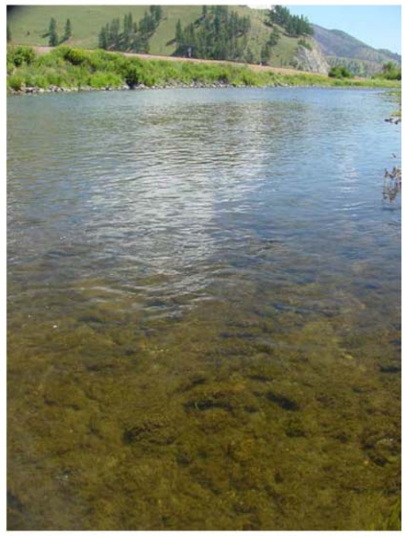

Photo C: 400 mg Chl a/m^2

Photo D: 1280 mg Chl a/m^2

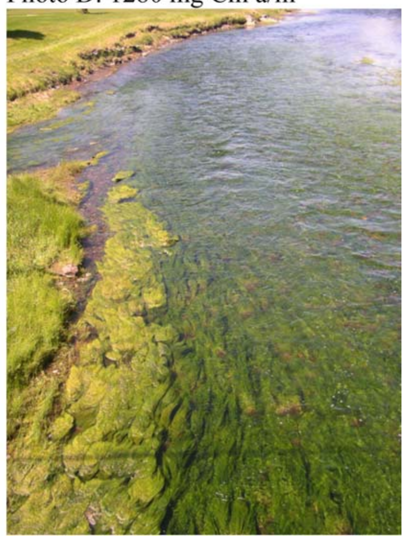

Photo E: 200 mg Chl a/m^2

Photo F: 150 mg Chl a/m^2

Photo G: 110 mg Chl a/m^2

Photo H: 300 mg Chl a/m^2

Acknowledgments
This project was funded by the Utah Division of Water Quality. Jakus was also funded by the Utah Agricultural Experiment Station; he is not affiliated with any Koch Foundation-funded entity at USU. Nelson is now a research scientist at the University of Montana Flathead Lake Biological Station. Supporting data may be obtained from the corresponding author. We thank Ben Holcomb, John Loomis, Mary Jo Kealy, and Nicholas von Stackelberg for helpful discussion throughout the project. We especially thank Michael Suplee for providing his original photographs and comments on an earlier version of this manuscript. The data and code needed to replicate our reported results are available at Jakus et al. (2017), https://doi.org/10.15142/T3NS7T. The views and opinions expressed in this manuscript do not necessarily reflect those of the Utah Division of Water Quality. All errors remain those of the authors.

References

Barica, J. (1981). Hypereutrophy—The ultimate stage of eutrophication. *Water Quality Bulletin*, *6*(4), 95.

Burkholder, J. M., Dickey, D. A., Kinder, C. A., Reed, R. E., Mallin, M. A., McIver, M. R., & Toms, D. (2006). Comprehensive trend analysis of nutrients and related variables in a large eutrophic estuary: A decadal study of anthropogenic and climatic influences. *Limnology and Oceanography*, *51*(1), 463–487.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., . . . Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, 1433–1436. https://doi.org/10.1126/science.aaf0918

Camfield, L., & Palmer-Jones, R. (2013). Three 'Rs' of econometrics: Repetition, reproduction, and replication. *Journal of Development Studies*, *49*(12), 1607–1614.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*, 181–210.

Dillman, D. (2000). *Mail and internet surveys: The tailored design method* (Vol. 2). New York, NY: John Wiley.

Dodds, W. K. (2006). Eutrophication and trophic state in rivers and streams. *Limnology and Oceanography*, *51*(12), 671–680. https://doi.org/10.4319/lo.2006.51.1_part_2.0671

Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., . . . Thornburgh, D. J. (2009). Eutrophication of U.S. freshwaters: Analysis of potential economic damages. *Environmental Science & Technology*, *43*(1), 12–19.

Dodds, W. K., & Welch, E. B. (2000). Establishing nutrient criteria in streams. *Journal of the North American Benthological Society*, *19*(1), 186–196.

Environmental Protection Agency. (2014). *Water quality standards handbook* (2nd ed., chap. 6, Rep. EPA-820-B-14-003). Washington, DC: Author. Retrieved from https://www.epa.gov/wqs-tech/water-quality-standards-handbook

Falconer, I. R. (1999). An overview of problems caused by toxic blue-green algae (cyanobacteria) in drinking and recreational water. *Environmental Toxicology*, *14*, 5–12.

Freese, J. (2007). Replication standards for quantitative social science: Why not sociology? *Sociological Methods and Research*, *36*(2), 153–172.

Greene, W. (2008). *Econometric analysis* (6th ed.). New York, NY: Macmillan.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Hamermesh, D. S. (2007). Viewpoint: Replication in economics. *Canadian Journal of Economics*, *40*(3), 715–733.

Herriges, J., Kling, C., Liu, C. C., & Tobias, J. (2010). What are the consequences of consequentiality? *Journal of Environmental Economics and Management*, *59*, 67–81.

Horner, R. R., Welch, E. B., & Veenstra, R. B. (1983). Development of nuisance periphytic algae in laboratory streams in relation to enrichment and velocity. In R.G. Wetzel (Ed.), *Periphyton of freshwater ecosystems* (pp. 121–131). The Hague, the Netherlands: Dr. W. Junk.

Jakus, P. M., Kealy, M. J., Loomis, J., Nelson, N., Ostermiller, J., Stanger, C., & von Stackelberg, N. (2013). *Economic benefits of nutrient reductions in Utah's waters*. Retrieved from https://deq.utah.gov/Pollutants/N/nutrients/studies/economic.htm

Jakus, P. M., Nelson, N., & Ostermiller, J. (2017). *Nutrient criterion data*. Logan, UT: Utah State University. https://doi.org/10.15142/T3Ns7T

Junker, B., & Buchecker, M. (2008). Aesthetic preferences versus ecological objectives in river restorations. *Landscape and Urban Planning*, *85*, 141–154.

Knowlton, M. F., & Jones, J. R. (2006). Natural variability in lakes and reservoirs should be recognized in setting nutrient criteria. *Lake and Reservoir Management*, *22*(2), 161–166.

Le Lay, Y.-F., Piegay, H., & Riveire-Honegger, A. (2013). Perception of braided river landscapes: Implications for public participation and sustainable management. *Journal of Environmental Management*, *119*, 1–12.

Lewis, K. E., Grebitus, C., & Nayga, R. M. (2016). U.S. consumers' preferences for imported and genetically modified sugar: Examining policy consequentiality in a choice experiment. *Journal of Behavioral and Experimental Economics*, *65*, 1–8.

Meitner, M. J. (2004). Scenic beauty of river views in the Grand Canyon: Relating perceptual judgments to locations. *Landscape and Urban Planning*, *68*, 3–13.

Nelson, N., Loomis, J., Jakus, P. M., Kealy, M. J., von Stackelberg, N., & Ostermiller, J. (2015). Linking ecological data and economics to estimate the total economic value of improving water quality by reducing nutrients. *Ecological Economics*, *118*, 1–9. https://doi.org/10.1016/j.ecolecon.2015.06.013

Paul, M. J., Walsh, B., Oliver, J., & Thomas, D. (2017). *Algal indicators in streams: A review of their application in water quality management of nutrient pollution* (Doc. 822B17002). Washington, DC: U.S. Environmental Protection Agency. Retrieved from https://www.epa.gov/sites/production/files/2017-06/documents/algal-indicators-whitepaper.pdf

Responsive Management. (2012). *West Virginia residents' opinions on and tolerance levels of algae in West Virginia Waters*. Harrisonburg, VA: Responsive Management. Retrieved from http://www.dep.wv.gov/WWE/Programs/wqs/Documents/WVAlgaeSurveReport_ResMgmt_WVDEP_2012.pdf

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Smith, A. J., Duffy, B. T., & Novak, M. A. (2015). Observer rating of recreational use of wadeable streams of New York State, USA: Implications for nutrient criteria development. *Water Research*, *69*, 195–209.

StataCorp. (2017). *Stata statistical software: Release 15*. College Station, TX: StataCorp LLC.

Stevenson, R. J., Bennett, B. J., Jordan, D. N., & French, R. D. (2012). Phosphorus regulates stream injury by filamentous green algae, DO, and pH with thresholds in responses. *Hydrobiologia*, *695*(1), 25–42.

Suplee, M. W., Watson, B., Teply, M., & McKee, H. (2009). How green is too green? Public opinion of what constitutes undesirable algae levels in streams. *Journal of the American Water Resources Association*, *45*(1), 123–140.

Utah County Health Department. (2016). *Utah county algal bloom advisories increased in 3 water bodies*. Retrieved from http://www.deq.utah.gov/NewsNotices/media/newsroom/docs/2016/08aug/Utah-County-Press-Release-Advisories-Increased-2016-08-22.pdf

Vossler, C. A., Doyon, M., & Randeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4), 145–171.