Statewide Stream/River Probabilistic Monitoring Network for the State of Oklahoma from 2008-2011

Final Report **September 3, 2013** FY-2011 WPCP-Monitoring Section 106 Supplemental, Project 2 (CA# I-00F432-01) –"The Statewide Stream/River Probabilistic Monitoring Network—Final Report"

Oklahoma Water Resources Board Water Quality Programs Division Monitoring and Assessment Section 3800 N. Classen, Oklahoma City, Oklahoma 73118 405-530-8800

Contact:

Monty Porter, Technical and Quality Manager, maporter@owrb.ok.gov Bill Cauthron, Monitoring Coordinator, wlcauthron@owrb.ok.gov

Technical assistance and advice provided by the United States Environmental Protection Agency Western Ecology Division. Special thanks to Tony Olsen, Dave Peck, and Tom Kincaid for invaluable assistance in both planning and analysis of data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

[Figure 12. Biological indicator condition extent estimated statewide for smaller streams and rivers](#page-34-0) [\(Strahler Order < 5\) from 2008-2011. Upper and lower bounds represent a 95% confidence](#page-34-0) interval. [...](#page-34-0) 35 [Figure 13. Biological indicator condition extent estimated statewide from 2008-2009. Upper and](#page-35-0) [lower bounds represent a 95% confidence interval...](#page-35-0) 36 Figure 14. Biological [indicator condition extent estimated statewide from 2010-2011. Upper and](#page-36-0) [lower bounds represent a 95% confidence interval...](#page-36-0) 37 [Figure 15. Stacked percentages of condition class estimates for](#page-38-0) study periods grouped by [stressors. \(Refer to Table 7 for stressor descriptions.\)...](#page-38-0) 39 [Figure 16. Stacked percentages of condition class estimates for stream size grouped by stressors.](#page-39-0) (Refer to Table 7 for stressor descriptions.) [..](#page-39-0) 40 [Figure 17. Stressor extent estimated statewide from 2008-2011. Upper and lower bounds](#page-40-0) [represent a 95% confidence interval. \(Refer to Table 7 for stressor descriptions.\)](#page-40-0) 41 [Figure 18. Stressor extent estimated statewide for larger streams and rivers \(Strahler Order > 4\)](#page-41-0) [from 2008-2011. Upper and lower bounds represent a 95% confidence](#page-41-0) interval. (Refer to [Table 7 for stressor descriptions.\)..](#page-41-0) 42 [Figure 19. Stressor extent estimated statewide for smaller streams and rivers \(Strahler Order < 5\)](#page-42-0) [from 2008-2011. Upper and lower bounds represent a 95% confidence interval. \(Refer to](#page-42-0) [Table 7 for stressor descriptions.\)..](#page-42-0) 43 [Figure 20. Stressor extent estimated statewide from 2008-2009. Upper and lower bounds](#page-43-0) [represent a 95% confidence interval. \(Refer to Table 7 for stressor descriptions.\)](#page-43-0) 44 [Figure 21. Stressor extent estimated statewide from 2010-2011. Upper and lower bounds](#page-44-0) [represent a 95% confidence interval. \(Refer to Table 7 for stressor descriptions.\)](#page-44-0) 45 [Figure 22. Sedimentation and other habitat stressors estimated statewide from 2008-2011. Upper](#page-45-0) [and lower bounds represent a 95% confidence interval. \(Refer to Table 7 for stressor](#page-45-0) descriptions.) [..](#page-45-0) 46 [Figure 23. Metal toxicity extent estimated statewide from 2010-2011. Upper and lower bounds](#page-46-0) [represent a 95% confidence interval. \(Refer to Table 7 for stressor descriptions.\)](#page-46-0) 47 [Figure 24. Relative risk of nutrient stressors affecting poor fish condition by study period.](#page-48-0) [\(upper/lower bounds represent a 95% confidence interval-CI\) \(* = significant relative risk-RR\)](#page-48-0) [..](#page-48-0) 49 [Figure 25. Relative risk of nutrient stressors affecting poor fish condition by waterbody size.](#page-48-1) [\(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-48-1) ... 49 [Figure 26. Relative risk of conductivity and turbidity stressors affecting poor fish condition by study](#page-49-0) [period. \(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-49-0) 50 [Figure 27. Relative risk of conductivity and turbidity stressors affecting poor fish condition by](#page-49-1) [waterbody size. \(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-49-1) 50 [Figure 28. Relative risk of sediment and habitat stressors affecting poor fish condition by waterbody](#page-50-0) [size. \(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-50-0) 51 [Figure 29. Relative risk of metal toxicity stressors affecting poor fish condition by waterbody size.](#page-50-1) [\(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-50-1) ... 51 [Figure 30. Relative risk of nutrient stressors affecting poor macroinvertebrate condition by study](#page-51-1) [period. \(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-51-1) 52 [Figure 31. Relative risk of nutrient stressors affecting poor macroinvertebrate condition by](#page-52-0) [waterbody size. \(upper/lower bounds represent a 95% CI\) \(* = significant RR\)](#page-52-0) 53 [Figure 32. Relative risk of conductivity and turbidity stressors affecting poor macroinvertebrate](#page-52-1) condition by study period. (upper/lower bounds represent a 95% CI) ($*$ = significant RR) ... 53 [Figure 33. Relative risk of conductivity and turbidity stressors affecting poor macroinvertebrate](#page-53-0) condition by waterbody size. (upper/lower bounds represent a 95% CI) ($*$ = significant RR)...54 [Figure 34. Relative risk of sediment and habitat stressors affecting poor macroinvertebrate condition](#page-53-1) [by waterbody size. \(upper/lower bounds represent a 95% CI\) \(* = significant RR\)..............](#page-53-1) 54

EXECUTIVE SUMMARY

Several agencies conduct water quality monitoring in the State of Oklahoma. These agencies meet complementary monitoring objectives that support the management of Oklahoma's surface waters. The two primary components of the statewide monitoring program include (a) the Beneficial Use Monitoring Program, a long-term, fixed-station water quality monitoring network of the Oklahoma Water Resources Board (OWRB), and (b) Oklahoma Conservation Commission's (OCC) Small-Watershed Rotating Basin Monitoring Program, targeting water quality and ecological conditions in waters flowing from 11-digit hydrologic units. The state recently completed a water quality monitoring strategy that describes their existing programs in detail and the monitoring objectives that cannot be met with existing resources (OWRB, 2012d). These objectives include the ability to make statistically valid inferences about environmental conditions throughout the state, based on a probabilistic selection of sites. Meeting this objective will improve the ability to make condition estimates required in section 305(b) of the Clean Water Act. This requirement includes a description of the quality of all lotic waters, and the extent that all waters provide for the protection and propagation of aquatic life. The Environmental Protection Agency (EPA) recently released guidance establishing the "10 Required Elements of a State Water Monitoring and Assessment Program" (USEPA, 2005). Among other things, the document states, "a State monitoring program will likely integrate several monitoring designs (e.g., fixed station, intensive and screening-level monitoring, rotating basin, judgmental and probability design) to meet the full range of decision needs. The State monitoring design should include probability-based networks (at the watershed or state-level) that support statistically valid inferences about the condition of all State water types, over time. EPA expects the State to use the most efficient combination of monitoring designs to meet its objectives."

From 2008-2011, Oklahoma completed its 2^{nd} and 3^{rd} statewide surveys of lotic waters. In SY 2008-2009, Oklahoma participated in the National Rivers and Streams Assessment (NRSA) and sampled fifty-two (52) stations equally proportioned across orders 1-4 and 5+, completing its second comprehensive survey. In SY 2010-2011, Oklahoma completed its third statewide probabilistic study with a sample size of 48 perennial streams and rivers. The new study population included perennial streams and rivers throughout Oklahoma, and continued through the NRSA draw into the remaining oversample sites. By combining the two studies, Oklahoma can report on several temporal scales, and on two (2) size classes—smaller and larger waterbodies. Temporal scales include:

- 52 sites in the 2008-2009 sampling period (NRSA study)
- 48 sites in the 2010-2011 sampling period (OWRB study)
- 100 sites over the 2008-2011 sampling period (combined study).

The probability-based survey was designed to assist Oklahoma's water quality managers in several ways. Furthermore, in keeping with the environmental goals of the state as outlined in the comprehensive water plan, an effective long-term management strategy based on sound science and defensible data can be developed using this data. The four over-arching goals were:

- 1. Estimate the condition of multi-assemblage biological indicators for Oklahoma's waters through a statistically-valid approach.
- 2. Estimate the extent of stressors that may be associated with biological condition.
- 3. Evaluate the relationship between stressors and condition for use in various long and short term environmental management strategies.
- 4. Assess waters for inclusion in Oklahoma's Integrated Water Quality Report.

To assess ecological and human health, one-time collections were made for a variety of biological, chemical, and physical parameters (Table 1). When sites were verified as target, a sampling schedule was implemented. All target sites were visited once (in rare instances twice) during a late spring to late summer index period (June 1 – August 30), under base flow conditions. The studies measured the condition of three biotic assemblages—fish, macroinvertebrates, and sestonic and benthic algae—and a variety of stressors, including nutrients, conductivity, turbidity, habitat and sedimentation, and toxics. Fish data were analyzed using two indices of biological integrity (IBI) commonly used in Oklahoma bioassessment studies, as well as the IBI developed by the NRSA. Macroinvertebrate data were analyzed using a Benthic-IBI (B-IBI) developed for Oklahoma benthic communities (OCC, 2005a) and commonly used by the OCC and OWRB Water Quality Divisions (OCC, 2008; OWRB, 2009 and 2010a; ODEQ, 2012), as well as the IBI developed by the NRSA. To estimate condition of algal biomass, chlorophyll-a concentrations were compared to several screening levels.

Data outputs include: 1) relative extent of indicator and stressor condition, 2) relative risk of stressors to indicators, and 3) attributable risk of stressors to indicator extent. Data will also combined with other sources and included in the 2014 303(d) assessment of the Oklahoma Integrated Water Quality Report.

Highlights of the relative extent include:

- For both fish and macroinvertebrates, nearly 35% of stream miles were classified in poor condition over the 4-year study period, and the poor category increased to greater than 40% from 2008-2009 and decreased to less than 25% from 2010-2011.
- When considering stream size, a greater percentage of large river stream miles are in poor condition than small streams.
- A relative small percentage of miles (10%) are classified in poor condition for benthic algae. a greater percentage of large rivers (22%) than small streams (6%) are in poor condition.
- For sestonic algae, the percentage of streams in poor condition across study years varies from nearly 20% (2008-2009) to nearly 30% from 2010-2011, while the percent in good condition is approximately 55% for all study periods, and approximately 60% of large river miles are in poor condition as compared less than 10% of small river miles.
- Phosphorus extent in poor condition is generally 30-40%, regardless of study period or source of screening limit, while the percent of total miles in good condition ranges from 40- 50%.
- Total nitrogen poor condition is from 25-40%.
- For conductivity, poor condition ranges from 10-22%, and is 40-55% in larger rivers, as opposed to 5% in small streams.
- For turbidity, poor condition is nearly 25%, and is 37% in larger rivers as opposed to 9% in smaller streams.
- Excess sedimentation from greater than 25% in streams to 35% in rivers, with poor condition ranging from 15% in 2008-2009 to greater than 50% from 2010-2011.

The current study allows for unique analysis between both study periods and waterbody size.

- For indicators, both fish and macroinvertebrates demonstrate a downward trend in poor condition between study periods, with only the fish having a significant downward trend.
- Conversely, both algal indicators show an upward trend, with only the benthic algae trend having significance.
- All but one of the total phosphorus stressors shows an upward trend between the two study periods, with only turbidity and sediment having a significant trend.

Attributable risk analyses provided the following results:

- Notably, for fish, elimination of sediment in large rivers could create a significant reduction of poor condition in fish as could reduction in conductivity.
- For macroinvertebrates, elimination of both total phosphorus and total nitrogen could have a significant effect on poor condition
- The elimination of phosphorus in small streams results in a nearly 14% lowering of the percent of miles in poor condition.
- As with fish, the elimination of conductivity is significant in some scenarios.
- Sestonic algal condition shows significant reduction in poor condition when turbidity, conductivity, and nutrients are eliminated.

Future study plans include the 2013-2014 National Rivers and Streams Assessment and a subsequent two-year statewide study beginning in 2015 (OWRB, 2013b). Substantive changes to the program will include

- Use of the NRSA protocols for large Wadeable and non-wadeable waterbodies.
- Use of NRSA habitat protocols for wadeable streams in concert with the current RBP habitat protocol.
- Inclusion of a second winter macroinvertebrate index period.
- Development of a periphyton taxonomic assemblage.
- Assessments at aggregated ecoregion scales used in the 2005-2007 assessment (OWRB, 2009)
- Change/trend analyses through the use of revisit sites.

INTRODUCTION

Several agencies conduct water quality monitoring in the State of Oklahoma. These agencies meet complementary monitoring objectives that support the management of Oklahoma's surface waters. The two primary components of the statewide monitoring program include (a) the Beneficial Use Monitoring Program, a long-term, fixed-station water quality monitoring network of the Oklahoma Water Resources Board (OWRB), and (b) Oklahoma Conservation Commission's (OCC) Small-Watershed Rotating Basin Monitoring Program, targeting water quality and ecological conditions in waters flowing from 11-digit hydrologic units. The state recently completed a water quality monitoring strategy that describes their existing programs in detail and the monitoring objectives that cannot be met with existing resources (OWRB, 2012d). These objectives include the ability to make statistically valid inferences about environmental conditions throughout the state, based on a probabilistic selection of sites. Meeting this objective will improve the ability to make condition estimates required in section 305(b) of the Clean Water Act. This requirement includes a description of the quality of all lotic waters, and the extent that all waters provide for the protection and propagation of aquatic life.

The Environmental Protection Agency (EPA) recently released guidance establishing the "10 Required Elements of a State Water Monitoring and Assessment Program" (USEPA, 2005). Among other things, the document states, "a State monitoring program will likely integrate several monitoring designs (e.g., fixed station, intensive and screening-level monitoring, rotating basin, judgmental and probability design) to meet the full range of decision needs. The State monitoring design should include probability-based networks (at the watershed or state-level) that support statistically valid inferences about the condition of all State water types, over time. EPA expects the State to use the most efficient combination of monitoring designs to meet its objectives." Until 2005, Oklahoma had several monitoring programs that met these requirements including the Beneficial Use Monitoring Program (BUMP) and the Rotating Basin Monitoring Program (RBMP) (OWRB, 20012d). Furthermore, the state has developed several programs to intensively monitor areas that have been listed on Oklahoma's 303(d) list of impaired waters (ODEQ, 2010).

In 2001, the state requested assistance with the design of a probabilistic approach to stream and river site selection from the U.S. Environmental Protection Agency, Office of Research and Development (ORD), Western Ecology Division (OWRB, 2006a). The study design was completed, but Oklahoma agencies remained unable to initiate further planning and implementation because of a lack of resources and commitment. In 2004, the OWRB and OCC took part in the National Wadeable Streams Assessment (WSA) (USEPA, 2006), which was fortuitous to future planning efforts for several reasons. First, the timing of the study coincided with discussions in the state about implementing a probabilistic design. Although money was a question, staff and management were worried staff time could not be spent performing all of the necessary reconnaissance work or sampling that is required in a random based monitoring program. Participating in the WSA instilled confidence that this type of monitoring could be accomplished without impeding the success of other programs. In fact, this facet of Oklahoma's monitoring program has only enhanced other programs. Second, because the state showed interest in implementing a random design, USEPA Region 6 began working with staff to find appropriate funding. The initial funding came through a Clean Water Act (CWA) Section 104(b)(3) grant. This money funded not only the initial year of study (2005), but an outcome was to investigate the feasibility of full implementation (OWRB, 2006a). The study investigated feasibility on two fronts—logistic and funding—finding that the logistic portion could be overcome through proper planning and coordination of staff. The funding, however, was not easily dealt with because of program priorities. In 2005, another funding opportunity came open when the USEPA announced further funding of the Regional Environmental Monitoring and Assessment Program (REMAP) (OWRB, 2009). Funding from the REMAP grant allowed the state to continue implementation of probabilistic monitoring for an additional two years through 2007. In that study, the OWRB completed a large-scale statewide assessment of perennial rivers and streams, as well as assessments for three large ecoregion groupings including the Western and High Plains, the Forested Plains and Flint Hills, and the Eastern Highlands. A significant limitation during that study was the inability to determine biological condition in large rivers.

In SY 2008-2009, Oklahoma participated in the National Rivers and Streams Assessment (NRSA) and sampled fifty-two (52) stations equally proportioned across orders 1-4 and 5+, completing its second comprehensive survey. In SY 2010-2011, Oklahoma completed its third statewide probabilistic study with a sample size of 48 perennial streams and rivers. The new study population included perennial streams and rivers throughout Oklahoma, and continued through the NRSA draw into the remaining oversample sites. By combining the two studies, Oklahoma can report on several temporal scales, and on two (2) size classes—smaller and larger waterbodies. Temporal scales include:

- 52 sites in the 2008-2009 sampling period (NRSA study)
- 48 sites in the 2010-2011 sampling period (OWRB study)
- 100 sites over the 2008-2011 sampling period (combined study)

The probability-based survey was designed to assist Oklahoma's water quality managers in several ways. Furthermore, in keeping with the environmental goals of the state as outlined in the Oklahoma Comprehensive Water Plan, an effective long-term management strategy based on sound science and defensible data can be developed using this data. The four over-arching goals were:

- 5. Estimate the condition of multi-assemblage biological indicators for Oklahoma's waters through a statistically-valid approach.
- 6. Estimate the extent of stressors that may be associated with biological condition.
- 7. Evaluate the relationship between stressors and condition for use in various long and short term environmental management strategies.
- 8. Assess waters for inclusion in Oklahoma's Integrated Water Quality Report.

The current assessment allows the state to make a statistically valid assessment of the condition of all of Oklahoma's streams/rivers, as required under Section 305(b) of the Clean Water Act (CWA) (ODEQ, 2012). The sample size allows for a statewide estimate of fish, macroinvertebrate, and algal condition on 3 temporal scales, as well as two size classes. Additionally, stressor extent is evaluated for a number of potential environmental stressors. Under the guidelines of the Integrated Listing Methodology (ODEQ, 2012), data allow for the assessment of the Fish & Wildlife Propagation beneficial use on more waters of the state. Although currently limited to certain beneficial uses and associated criteria, the support status of more waters can be determined. Future work may allow for more comprehensive 303(d) assessments so that the support status of probabilistic sites may be fully vetted. Finally, the survey provides information that will allow for better long- and short-range planning and resource allocation. A benefit of probabilistic design is that data results can be applied in a much broader context. For example, the relationship of condition can be associated with stressor extent through methodologies like relative risk analysis. The current study yields a wealth of biological, chemical, and physical data across a broad gradient of environmental conditions, supporting evaluation of these indicator relationships. Data can be used to calibrate existing biocriteria ranges, establish reference condition, and assist in nutrient criteria development. When integrated with fixed-station networks, it can assist in identifying local areas of concern. Also, although not accomplished by this report, landscape metrics can be associated with stressors and condition to develop predictive models. Probabilistic data can assist in efforts to regionalize environmental concerns. A bottom up approach to management identifies not only statewide issues but allows managers to identify local and regional concerns first, which often lead to issues farther down the watershed, and put resources where they are needed. The probabilistic methodology adds a valuable layer to that management approach.

METHODS

Study Design

An unequal probability random tessellation stratified (RTS) survey design (Stevens 1997, Stevens and Olsen 2004) was used to select stream sample sites across the state (USEPA, 2012 and Appendix D-1). The original design for the 4-year study emanated from Oklahoma's site file for the 2008-2009 NRSA. Unequal probability categories were defined separately for wadeable streams (1st to 4th order) and non-wadeable rivers (5th to 10th order). The terms wadeable and non-wadeable were used to designate Strahler order classes and did not imply that the streams were actually wadeable or non-wadeable, as defined by protocol. For the wadeable stream category, unequal selection probabilities were defined for 1st, 2^{nd} , 3^{rd} , and 4^{th} order streams so that an equal number of sites would occur for each order. Then these unequal selection probabilities were adjusted by the Wadeable Streams Assessment (WSA) nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category. For the nonwadeable river category, unequal selection probabilities were defined for 5th, 6th, 7th, and 8th + Strahler order Rivers so that the expected number of sites nationally would be 350, 275, 175, and 100 sites, respectively. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category. Additionally, certain sites were selected as revisit sites from the 2004 Wadeable Streams, and included in the initial study design, weighted equally across the Strahler order categories mentioned above. In Oklahoma for the 4-year study period, the expected sample size was 51 for both wadeable streams and non-wadeable rivers. Oversample sites were provided for each Strahler order grouping. Site replacement was done within the two major Strahler order categories, 1st-4th and $5th +$.

The study was spatially, temporally and hydrologically limited. Spatially, the study was limited to only streams defined as perennial in flow and excluded all sites within a reservoir flood pool. Temporal limitations were defined by biological index periods. The index period for the fish assemblage in Oklahoma was May 15th through September 15th with an optional extension to October 1st if the stream had not risen above summer seasonal base flow (OWRB, 2010b). The index habitat period for the macroinvertebrate assemblage in Oklahoma was June $1st$ through August $30th$ with collections completed in as short a time period as possible (OWRB, 2010c). Hydrologically, the study was limited by both an extended drought in SY-2011 as well as excessive rains and flooding in SY-2008. This impeded study progress in several ways. Sites originally verified as target sites were removed and an oversample site visited because of site changes between the period of reconnaissance and sampling. Additionally, several sites had partial collections because conditions changed between the period of macroinvertebrate/water sampling and fish sampling, or vice-versa. Furthermore, all of the smaller Strahler order category sites were ultimately evaluated. Because of accessibility issues related to drought in SY-11, only 48 sites were available for inclusion in the 2010-2011 study.

The study and subsequent site selection were designed to allow for three reporting periods and subcategorization of "small" and "large" sites. The reporting periods include 2008-2011 (n = 100), 2008- 2009 (n = 52), and 2010-2011 (n = 48). The 2008-2011 was sub-categorized to evaluate small (1st-4th Strahler Order) and large (5th+ Strahler Order) waterbodies. For each subcategory, an "n" of 50 was achieved. The oversample sites from the original NRSA sample design were used to provide sites for the 2010-2011 study.

Site Reconnaissance

Limited accessibility is the most serious problem with any probabilistic study. Unlike a fixed station design, study sites are typically not accessible by public roads and may only be accessed by foot. Compounding the problem is private ownership of land and the need to respect a landowner's choice of who may or may not access the property. Finally, probabilistic sites are selected from data frames that are not 100% accurate and may include non-candidate sites. Fortunately, proper planning and having an excess of available oversample sites can alleviate these issues. During the EPA's Wadeable Streams Assessment (USEPA, 2006) and Oklahoma's 1st Statewide Probabilistic Study (REMAP) (OWRB, 2009), the OWRB developed (with assistance from EPA documentation) and implemented a three-stage reconnaissance plan.

The first stage of planning was a "desk top" reconnaissance to determine if the proposed site was a candidate site. Candidate sites must meet certain criteria, including: 1) perennial flow, 2) not within normal pool elevation of a lake (oxbows or reservoirs), 3) not a wetland/swamp dominated river, 4) accessible by foot, and 5) landowner permission granted. Initially, each site was located using a variety of resources including topographic maps (OWRB, 2011), and other GIS mapping tools (NACEC, 1997). For each site, a site reconnaissance and tracking form (Figure 3) was created with the ultimate determination made to "accept" or "reject". At the outset, required hydrological characteristics were verified, and if not met, the site was rejected without further consideration. Then, a series of site maps containing at least two geographic scales were included with the site tracking form, and the necessary information to determine landowner was collected, including legal description of site and county. County assessor offices were the main source of landowner information. However, for some problem sites, staff used a variety of other resources including development of relationships with local realtors/developers or personal visits to nearby residences. Finally, a landowner permission packet was sent to each landowner, including a standardized permission letter (Figure 4), maps, a study brochure, and self addressed/stamped envelope for them to review and mail back to the OWRB either approving or disallowing access to their property. Based on landowner response, the site was accepted, accepted with restrictions/further instructions, or rejected. However, even when good landowner information was available, response to permission requests was occasionally slow for a variety of reasons, and therefore, a two stage process was developed to deal with slow responses. After two to three weeks, staff attempted contact by phone, and if unsuccessful, would send a reminder postcard. If still unsuccessful, inperson contact was attempted. If each of these attempts failed, the site was rejected.

Once site accessibility was verified (i.e., accepted) and a site was labeled as a study target site, a second planning stage was initiated. The planning objective was simply to collect thorough, welldocumented information to assist field crews in locating and accessing the sampling reach. Because of color aerial satellite imagery, much of this information was gathered from the desktop. Notes were made and included in the tracking form of special considerations including hazards, best route of entry, time of travel, etc. Unfortunately, some sites required an on-site initial visit to complete the planning phase. Concerns did arise about the cost versus benefit of an extra site visit. However, over the course of three years, crews discovered that much of the information collected during the initial on-site planning visit was of great benefit on the actual day of sampling. Furthermore, because sites could be visited in batches and only one staff member was required, little expense was incurred.

The final planning stage involved all activities up to the first sampling visit, and involved compiling a complete site packet. The packet incorporated all information gathered in stages one and two,

including a completed tracking form, landowner permission letter, and pertinent pictures and maps. In addition, all necessary field forms and labels were compiled and a checklist of equipment needed was completed.

Figure 1. Template site reconnaissance and tracking form used during study.

Figure 2. Template landowner permission letter used during study.

Data Collection

To assess ecological and human health, one-time collections were made for a variety of biological, chemical, and physical parameters (Table 1). When sites were verified as target, a sampling schedule was implemented. All target sites were visited once (in rare instances twice) during a late spring to late summer index period (June 1 – August 30), under base flow conditions. Collections included a comprehensive water chemistry sample and measurement of *in situ* water quality parameters, including water temperature, dissolved oxygen, pH, specific conductance, and turbidity. Additionally, biological assemblages were collected, including fish, macroinvertebrates, phytoplankton, and benthic periphyton. A comprehensive suite of physical habitat, riparian and human health influence measurements were made, as well as a variety of site observational information. In the event that a full collection could not be completed during the index period, an additional collection may have occurred for fish after May 10 or before October 15. Depending on circumstances, information was collected during the same site visit. Additionally, a winter index period was added for macroinvertebrates and water chemistry during the 2010 and 2011 sample years.

Table 1. Water quality variables included in study.

From 2008-2009, all collections strictly followed the NRSA field operations manual (USEPA, 2009a) and Quality Assurance Project Plan (USEPA, 2009b). Sample analyses for these years were provided by the NRSA contract laboratories and data/assessments for all samples and assemblages were provided by the USEPA through either their National Aquatic Resource Survey (NARS) sharefile portal [\(https://nars.sharefile.com/\)](https://nars.sharefile.com/) (USEPA, 2012) or personal communication from EPA staff (Mitchell, 2013).

For study years 2010-2011, data for water quality variables was collected in one of two ways (OWRB, 2010e). Several variables (pH, dissolved oxygen, water temperature, and specific conductance) were monitored *in-situ* utilizing a Hydrolab® Minisonde or YSI® multi-probe instrument or with single parameter probes. Regardless of instrumentation and in accordance with manufacturer's specifications and/or published SOP's, all instruments (except water temperature) were calibrated at least weekly and verified daily with appropriate standards. The measurement was taken at the deepest point of the channel at a depth of at least 0.1 meters and no greater than onehalf of the total depth. The data were uploaded from the instrument and saved to a data recorder, transferred manually to a field log sheet, and manually entered into the OWRB Water Quality database. Data for all other variables were amassed from water quality samples collected at the station. Grab samples were collected by one of two methods—a grab or a composite grab. The most common method employed was a grab sample, which was used in streams with a single, wellmixed channel. The sample was collected at the deepest, fastest flowing portion of the horizontal transect by completely submerging the bottle, allowing it to fill to the top, and capping the bottle underwater. Composite grabs were collected in rivers with multiple channels and were aliquotted into sample bottles using a clean splitter-churn. Each sample included three bottles for general chemistry analyses (two ice preserved and one sulfuric acid preserved), one bottle for metals analysis (nitric acid preserved), and one bottle each for field chemistry analysis and sestonic chlorophyll-a (ice preserved and kept dark). For benthic chlorophyll-a, a sample was composited, placed on ice to be preserved, and kept dark. The Oklahoma Department of Environmental Quality-State Environmental Laboratory (ODEQ-SEL) in accordance with the ODEQ's Quality Management Plan (QTRACK No. 00-182) (ODEQ, 2007) analyzed samples for most parameters listed in Table 4. OWRB personnel measured nitrogen and ortho-phosphorus using Hach® colorimeter protocols, hardness and alkalinity using Hach $^\circ$ titration protocols, and nephelometric turbidity using a Hach $^\circ$ Portable turbidometer.

Samples for algal biomass were collected in both the sestonic and benthic zones of each waterbody and processed in accordance with standard procedures outlined (OWRB, 2006b). Sestonic, or water column, samples were processed from water collected during the general water quality collection. A benthic sample was processed from a reach-wide composite. Benthic filters were extracted using an alternate method, whereby filters are placed in a standard aliquot of ethanol (25 mL) and extracted at room temperature for at least 72 hours. All chlorophyll-a samples were analyzed by the ODEQ-SEL under the previously mentioned QMP (ODEQ, 2007). Additionally, a 50-mL sample was collected from both the water column and the benthic composites for subsequent sestonic and benthic algal ID analysis. Samples were preserved with 10% formalin, wrapped with foil, and placed at 4° C.

Biological assemblages included aquatic macroinvertebrates and fish that were collected in accordance with Oklahoma's Rapid Bioassessment Protocols (RBP) (OWRB, 1999) and the OWRB's biological collection protocols (OWRB, 2010b and 2010d). Collections were completed over a 150-4000 meter reach depending on wetted width. Fish were collected during the summer index period using a pram or boat electrofishing unit depending on wadeability. The pram unit consisted of a Smith-Root 2.5 generator powered pulsator (GPP) attached to a 3000W Honda generator, and were operated with AC output current at 2-6 amps. The boat was equipped with a 9.0 GPP powered by a 9,000 Kohler generator, and operated at an AC output range of 7-20 amps. A battery powered Smith-Root backpack generator was used on rare occasions in sites with less than 1-meter average wetted width. Using two netters with ¼ inch mesh dipnets, collections were made in an upstream direction with target effort depending on reach length, site conditions, and protocol. When existing habitats existed could not be effectively electrofished, supplemental or stand-alone collections were made using 6' X 10-20' seines of $\frac{1}{4}$ inch mesh equipped with 8' brailles. Fish were processed at several intervals during each collection. The majority of fish were processed in the field, including enumeration and identification to species. Representative site voucher collections were made with a combination of appropriate photodocumentation and

representative species vouchers. Fish that were not readily identifiable were fixed in 10% formalin and returned to the OWRB laboratory for identification and enumeration. Additionally, all representative voucher fish were fixed in a 10% formalin solution, subsequently preserved in 80% ethanol, and, along with photodocumentation, permanently housed in the OWRB fish collection library.

Aquatic macroinvertebrate collections were made during the summer and winter index period of each study year (OWRB, 2010d). Each sampling event included a variety of samples as defined in the OWRB's macroinvertebrate collection protocols. At wadeable sites, staff collected samples from available targeted habitats, including streamside vegetation, woody debris, and rocky riffles. The streamside vegetation and woody debris collections were semi-qualitative samples collected over flowing portions of the reach for total collection times of three and five minutes, respectively. The streamside sample was collected using a 500-micron D-frame net to agitate various types of fine structure sample including fine roots, algae, and emergent and overhanging vegetation. Likewise, the wood sample was collected using a 500-micron D-frame net to agitate, scrape, and brush wood of any size in various states of decay. Additionally, wood that could be removed from the stream was scanned for additional organisms outside the 5-minute sampling time. The riffle collection was a quantitative sample compositing three kicks representing slow, medium and fast velocity rocky riffles within the reach. Each sub-sample was collected by fully kicking one square meter into a 500 micron Zo seine. At non-wadeable sites, a large river collection protocol was used, with the subprotocol determined by the dominant reach substrate, either fine or coarse substrate. In each protocol, the dominant substrate is sampled at each transect, and within each sub-reach, the dominant targeted habitat is sampled. The primary difference between the sub-protocols was the treatment of samples. The coarse protocol requires that all samples are processed and composited in a final collection type called large coarse-composite (LRC-Comp). While at the large river fine (LRF) sites, collections were kept separate and processed as LRF-THab (targeted habitat) and LRF-Sub (substrate) samples. At all LR sites, a riffle composite is collected, if available. All samples were field post-processed in a 500-micron sieve bucket to remove large material and silt in an effort to reduce sample size to fill no more than 34 of a quart sample jar. Additionally, all nets and buckets were thoroughly scanned to ensure that no organisms were lost. After processing, each sample type was preserved independently in quart wide mouth polypropylene jars with ethanol and interior and exterior labels were added. Prior to taxonomic analysis, all samples were laboratory processed by study personnel to obtain a representative 100 and 300-count subsample, with a large/rare scan (OWRB, 2010d). After sorting, the subsamples were sent to the contract laboratory of record for identification and enumeration. Taxonomic data for each sample were grouped and metrics calculated by the contract laboratory. In general, most organisms were identified to genera with midges identified to tribe. The two contract laboratories used in the study were Environmental Services and Consulting (Lynchburg, VA) and Rhithron Associates (Missoula, MT).

Additionally, a detailed habitat assessment was made targeting in-stream substrate, habitat, width and depth, bank and riparian measurements, and human disturbance characteristics. The collections included both Oklahoma's semi-qualitative RBP habitat protocols (OWRB, 2010c), and the NRSA semi-quantitative habitat protocols (USEPA, 2009a). To date, the USEPA assessments have not been processed.

Discharge and/or stage data were also collected at each station (OWRB, 2005). Flow was determined through several methods including direct measurement of instantaneous discharge using a flow meter, interpolation of flow from a stage/discharge rating curve developed by the United States Geological Survey (USGS) or the OWRB, or through estimation of discharge using a float test (OWRB, 2004).

For a more detailed discussion of sampling procedures, please contact the OWRB/Water Quality Programs Division at (405) 530-8800 for copy of the BUMP Standard Operating Procedures (SOP) or visit the OWRB website at [http://www.owrb.state.ok.us/quality/monitoring/monitoring.php#SOPs.](http://www.owrb.state.ok.us/quality/monitoring/monitoring.php#SOPs)

Analytical Methods

Condition classes for biotic assemblages and stressors were assigned by either the USEPA or OWRB, depending on study year. All data collected from 2008-2009 were processed and assessed by USEPA staff, excluding wadeable fish and chlorophyll-a data. All data collected from 2010-2011, as well as chlorophyll-a data from 2008-2009, were processed and assessed by OWRB staff.

Analysis of Fish Biological Condition.

Fish data were analyzed using two indices of biological integrity (IBI) commonly used in Oklahoma bioassessment studies, as well as the IBI developed by the NRSA. State biocriteria methods are outlined in Oklahoma's Use Support Assessment Protocols (OWRB, 20012b). In addition, an IBI commonly used by the OCC's Water Quality Division was used to provide an alternative bioassessment (OCC, 2005a and 2008; ODEQ, 2012). All metrics and IBI calculations were made using the OWRB's "Fish Assessment Workbook", an automated calculator OWRB staff built in Microsoft Excel (OWRB, 2012a). The NRSA condition assessments were taken from the tabular fish condition file on the USEPA's NARS sharefile site (USEPA, 2012). The multi-metric index (MMI) developed by the NRSA is described in Appendix D-3.

Oklahoma's biocriteria methodology (OKFIBI) uses a common set of metrics throughout the state (Table 2). Each metric is scored a 5, 3, or 1 depending on the calculated value, and scores are summed to reach two subcategory totals for sample composition and fish condition (OWRB, 2012b). The two subcategories are then summed for a final IBI score. The score is compared to ecoregion biocriteria to determine support status. For example, if the final IBI score is between 25-34, the status for sites in the Ouachita Mountain Ecoregion is deemed undetermined. Likewise, for scores greater than 34 and less than 25, the status is supported or not supported, respectively.

The OCCFIBI uses "a modified version of Karr's Index of Biotic Integrity (IBI) as adapted from Plafkin et al., 1989" (OCC, 2008; ODEQ, 2012). The metrics as well as the scoring system are in Table 3. Metric scores are calculated in two ways for both the test site and composite reference metric values of high-quality streams in the ecoregion (OCC 2005). Species richness values (total, sensitive benthic, sunfish, and intolerant) are compared to composite reference value to obtain a "percent of reference". A score of 5, 3, or 1 is then given the site depending on the percentages outlined in Table 6, while the reference composite is given a default score of 5. Proportional metrics (% individuals as tolerant, insectivorous cyprinids, and lithophilic spawners) are scored by comparing the base metric score for both the test site and the reference composite to the percentile ranges given in Table 3. After all metrics are scored, total scores are calculated for the test and composite reference sites. Finally, the site final score is compared to the composite reference final score and a percent of reference is obtained. The percent of reference is compared to the percentages in Table 4 and an integrity classification is assigned with scores falling between assessment ranges classified in the closest scoring group.

Fish taxonomic results for each site were analyzed to produce a raw score for the OKFIBI and a percent of reference score for the OCCFIBI. Additionally, when available, the condition class determined from the NRSA analysis was included in the evaluation. A preponderance of these assessments were used to then assign a final condition class of good, fair, and poor for each of the 3 study periods, as well as large and small streams.

Table 2. Index of biological integrity used to calculate scores for Oklahoma's biocriteria. Referenced figures may be found in OAC 785:15: Appendix C (OWRB, 2012b).

Table 3. Metrics and scoring criteria used in the calculation of OCC's index of biological integrity (OCC, 2008; ODEQ, 2012).

Table 4. Integrity classification scores and descriptions used with OCC's index of biological integrity (OCC, 2008; ODEQ, 2012).

Analysis of Macroinvertebrate Biological Condition

Macroinvertebrate data were analyzed using a Benthic-IBI (B-IBI) developed for Oklahoma benthic communities (OCC, 2005a) and commonly used by the OCC and OWRB Water Quality Division (OCC, 2008; OWRB, 2009 and 2010a; ODEQ, 2012), as well as the IBI developed by the NRSA. The metrics and scoring criteria (Table 5) are taken from the original "Rapid Bioassessment Protocols for Use in Streams and Rivers" (Plafkin et al., 1989) with slight modifications to the EPT/Total and Shannon-Weaver tolerance metrics (OCC, 2008). Metrics were calculated by OWRB contractors and IBI calculations were made using the OWRB's "B-IBI Assessment Workbook v. 3.0", an automated calculator built by OWRB Staff in Microsoft Excel (OWRB, 2012a). The NRSA condition assessments were taken from the tabular macroinvertebrate condition file on the USEPA's NARS sharefile site (USEPA, 2012). The IBI developed by the NRSA is described in Appendix D-4.

Calculation of the B-IBI is similar to the fish OCC-IBI discussed previously. Metric scores are calculated in two ways for both the test site and the composite reference metric values of highquality streams in each ecoregion (OCC, 2008). Species richness (total and EPT) and modified HBI values are compared to the composite reference value to obtain a "percent of reference". A score of 6, 4, 2 or 0 is then given the site depending on the percentages outlined in Table 5, while the reference composite is given a default score of 6. Proportional metrics (% dominant 2 taxa and %EPT of total) as well as the Shannon-Weaver Diversity Index are scored by comparing the base metric score for both the test site and the reference composite to the percentile ranges given in Table 5. After all metrics are scored, total scores are calculated for the test and composite reference sites. The site final score is then compared to the composite reference final score and a percent of reference is obtained. The percent of reference is compared to the percentages in Table 6 and an integrity classification is assigned with scores falling between assessment ranges classified in the closest scoring group.

Macroinvertebrate taxonomic results for each site were analyzed to produce a percent of reference score for the OKBIBI. From these scores, biological integrity classifications were assigned. For NRSA sites, the condition classification assigned by the NRSA was used because the samples were processed as 500 individual sub-samples. Instead of rarifying samples to a 100 individual subsample to allow use in Oklahoma's B-IBI, it was decided that using NRSA condition assignments was more defensible and efficacious for final data analyses. Furthermore, the NRSA IBI was used to assign condition classes for large rivers that were too large to be processed through Oklahoma B-IBI. These samples were compared to national reference metrics and screening limits developed for the NRSA.

Table 5. Metrics and scoring criteria used in the calculation of the B-IBI (OCC, 2008; ODEQ, 2012).

Table 6. Integrity classification scores and descriptions used with the B-IBI (OCC, 2008; ODEQ 2012).

Analysis of Algal Biomass

Algae are important in aquatic ecology acting as an important primary producer in aquatic food webs providing a food source for a wide variety of fish and macroinvertebrates. Furthermore, algae are indispensable producers of oxygen for aquatic organisms. However, algal blooms are also an important indicator of water quality perturbance and nutrient productivity. Introduction of nutrients to waterbodies occurs through a number of sources including runoff from urban and agricultural areas, wastewater treatment discharges, and a variety of other sources. As nutrient concentrations increase, uptake by primary producers increases and leads to algal blooms, as well as an increased standing crop. As eutrophication happens, aquatic life and human health beneficial uses can become impaired, as well as the aesthetic and recreational appeal of waterbodies being drastically reduced.

In order to quantify eutrophication, algal biomass was measured in both the benthic (i.e., periphyton) and water column (i.e., sestonic) areas of all study streams. Various measures exist to determine algal biomass including chlorophyll-a and ash free dry mass. For this study, chlorophyll-a concentrations were calculated because the Oklahoma Water Quality Standards (OWQS) (OWRB, 2012c) provides screening levels for both periphyton and sestonic chlorophyll-a.

To estimate condition of algal biomass, chlorophyll-a concentrations were compared to several screening levels. For benthic chlorophyll-a, several screening levels were used. First, Oklahoma's Use Support Assessment Protocol (USAP) (OWRB, 2012b) provides a screening level for periphyton chlorophyll-a in the aesthetic beneficial use. A value of 100 mg/m² represents a nuisance level for periphyton algae, and was used as the cut-point for poor-fair condition. Second, the OWRB has collected periphyton chlorophyll-a across the state for several programs throughout the years. To provide an alternate screening level, the $25th$ percentile of all OWRB benthic data were calculated at 45.7 mg/m², which was used as the cut-point for fair-good condition. Similarly, several screening levels were established for sestonic chlorophyll-a. The OWQS- includes a standard for sensitive water supplies of 10 mg/m³ (SesChI10) of chlorophyll-a (OWRB, 2012c), which was set as the fair-good cut-point for condition assessment. Additionally, to establish the cutpoint for the poor-fair condition, the distribution of all OWRB sestonic chlorophyll-a data were considered as a screening level (OWRB, 2009). The mean of all concentrations calculates at 19 mg/m³ and was set as the poor-fair cut-point for sestonic chlorophyll-a analyses.

Stressor Methodology

During each visit a number of physical and water quality parameters were collected. These included nutrients, *in situ* measurements, metals, and salinity. Each of these may have some effect on the conditions analyzed in the previous results section. This effect can lead to decreased biological integrity (e.g., the effect of nutrients on fish condition) or may be responsible for the increase in a negative condition (e.g., the effect of total phosphorus on algal biomass concentration). Quantifying stressor extent is important for a variety of reasons including development and refinement of water quality screening levels and criteria, location of hotspots, and understanding the cause and effect relationship between stressors and indicators of biological integrity and human health concerns. Stressor descriptions are given in Table 7. The final stressor methodology for chemistry is detailed in Appendix D-5.

Table 7. Descriptions of stressors affecting biological condition.

Nutrient stressors include measures of total phosphorus and total nitrogen (nitrate + nitrite + total Kjeldahl nitrogen). For comparison, two sources were used to determine screening levels for each parameter giving a variety of nutrient levels based upon stream characteristics and/or regional variation (Table 7). First, regional nutrient criteria were developed based on Omernik Level III ecoregions. The lower ender thresholds represent the 25th percentile of data from a variety of sources (USEPA, 2000a, 2000b, 2001a, 2001b; OWRB, 2009), while the upper end thresholds were developed from OCC regional monitoring data (OCC, 2005b, 2006a, 2006b, 2007, 2008). Second, the NRSA developed nutrient thresholds at a Level II ecoregion scale as described in Appendix D-5. The nutrient cut-point thresholds are in Table 8.

Additionally, both salinity and turbidity were evaluated as water quality stressors and are described in Table 7. Conductivity was used as a surrogate for salinity and several sources including both the USEPA regional criteria development (USEPA, 2000a, 2000b, 2001a, 2001b) and regional screening limits developed for Oklahoma's original statewide assessment (OWRB, 2009). Turbidity screening levels were only based on the USEPA regional criteria development reports. The cutpoints for conductivity and turbidity are provided in Table 9.

Numerical criteria for metals are housed in Appendix G, Table 2 of the OWQS (OWRB, 2012c). The OWQS provides criteria for a number of metals but only cadmium, copper, lead, selenium, and zinc are considered in this study. These analytes have both ecological and human health significance and appear more regularly in Oklahoma's Integrated Report as causes of impairment (ODEQ, 2012c). No other metals showed any level of potential impairment in the study. To facilitate analysis, dissolved metals concentrations were compared to dissolved chronic criterion.

Sedimentation was analyzed as a potential stressor to biological condition by using a combination of the state rule and NRSA condition assessments. For sites monitored as part of the NRSA, the sedimentation assessments were taken from the tabular habitat condition file on the USEPA's NARS sharefile site (USEPA, 2012), and the NARS methodology is described in Appendix D-2. For sites monitored in 2010-2011, metrics were calculated based on results from Oklahoma's Rapid Bioassessment Protocol (OWRB, 1999, 2010c, 2012b). The assessment consists of a variety of measures including flow, stream width and depth, substrates, embeddedness, habitat classification (i.e., pool, run, and riffle), fish cover, presence of point bars, erosion, and riparian structure. Metrics are scored based on predetermined ranges and a total score is obtained. Oklahoma's USAP (OWRB, 2012b) contains a protocol for determining sedimentation based upon loose bottom substrates (%LBS), embeddedness (%Emb), and presence of deep pools (%DP). Screening levels for sedimentation metrics are determined by comparing final site scores to a percent of reference condition. The reference condition is derived from the habitat scores for ecoregion based high quality sites developed by the OCC (2005a). For the most part, all high quality sites in an Omernik Level III ecoregion were used to develop reference condition. However, in certain ecoregions, some Omernik Level IV ecoregions were broken out from the whole. Omernik Level IV ecoregions used are the Broken Red Plains and Cross Timbers Transition of the Central Great Plains and the Arbuckle Uplift of the Cross Timbers. Additionally, the reference condition used is separated by aquatic life tier, and sites used to determine reference condition are required to be within 2 Strahler orders of the test stream. Finally, the cut-points for poor-fair-good are based on pre-determined percent of reference for each metric, with 2 or 3 metrics deemed to be fair or poor, respectively. Additionally, both instream cover and riparian vegetative cover were also evaluated as part of the NRSA. These stressors are included in the analysis of NRSA sites.

Statistical Methods

The processing of data for relative extent, relative risk, and attributable risk values were accomplished with R-statistical Software (R Foundation, 2013) using R-scripts developed for the NARS program (Van Sickle, 2012). Adjusted site weights were calculated and provided by the USEPA (Kincaid, 2013). Other analyses were performed using Minitab statistical software (Minitab, 2013). References to ecoregions throughout this document refer to those published by USEPA (Omernik, 1987; Woods et al., 2005).

	TN NRSA	TN NRSA	TN ECO	TN ECO	TP NRSA	TP NRSA	TP ECO	TP ECO
	Poor_Fair	Fair Good	Poor Fair	Fair Good	Poor Fair	Fair Good	Poor Fair	Fair Good
Ecoregion	(mg/L)							
Southwest Tablelands	1.570	0.698	1.050	0.450	0.095	0.052	0.055	0.025
Central Great Plains	1.570	0.698	1.600	0.840	0.095	0.052	0.130	0.090
Cross Timbers	1.570	0.698	0.900	0.680	0.095	0.052	0.110	0.038
Arbuckle Uplift	1.570	0.698	1.500	0.680	0.095	0.052	0.050	0.038
South Central Plains	2.078	1.092	0.750	0.385	0.108	0.056	0.070	0.050
Ouachita Mountains	0.535	0.296	0.450	0.300	0.024	0.018	0.025	0.010
Arkansas Valley	0.535	0.296	0.683	0.270	0.024	0.018	0.060	0.043
Ozark Highlands	0.535	0.296	1.500	0.379	0.024	0.018	0.070	0.007
Central Irregular Plains	3.210	1.750	1.150	0.712	0.338	0.165	0.160	0.093

Table 8. Ecoregion screening levels used as good/fair/poor cut-points for nutrient stressor analyses (Appendix D-5) (OWRB, 2009).

Table 9. Ecoregion screening levels used as good/fair/poor cut-points for conductivity and turbidity stressor analyses. (Appendix D-5) (OWRB, 2009)

RESULTS—EXTENT AND CONDITION ESTIMATES

Site Evaluation

For the study, a total of 177 randomly chosen sites were evaluated as candidate target sites, representing a total of 36,003 stream miles. Stream miles determined to be target, or sampleable, varied per study period (Figures 3-5). The total sampleable stream miles assessed per study period breaks down as follows:

- 21,019 miles for study period 2008-2011
- 25,466 miles for study period 2008-2009
- 15,572 miles for study period 2010-2011

The dramatic variation between the initial and subsequent 2-year study periods is obviously the number of rejected sites during the evaluation process. Although access denials increased between the study periods, the percentage of dry stream miles evaluated increased by over 300% from 3,094 to 10,605 evaluated miles, accounting for the dramatic decrease in assessed stream miles from reporting period to reporting period. Inaccessible and impounded miles were nearly equivalent across study periods. Furthermore, Figures 6 and 7 show a breakdown between large and small streams. Notably, small stream miles outnumbered large stream miles nearly 3.5:1, the majority of accessibility issues occurred in small streams.

Figure 3. Site evaluation status for study period 2008-2009 (total miles = 36,003).

Figure 4. Site evaluation status for study period 2010-2011 (total miles = 36,003).

Figure 5. Site evaluation status for study period 2008-2011 (total miles = 36,003).

Figure 6. Site evaluation status for small streams from 2008-2011 (total miles = 27,494).

Figure 7. Site evaluation status for large streams from 2008-2011 (total miles = 8,509).

Biological Indicator Condition Extent

Statewide condition extent estimates were made for benthic macroinvertebrates, fish, phytoplankton (sestonic algae) at two levels, and periphyton. For each biotic assemblage, the indicator condition was categorized as good, fair, or poor based on methodology described in the "Methods" section, and percentages for each condition category are based on "percent of total miles". In Figures 8-9, good/fair/poor estimates are grouped for each indicator by both study periods and size. In Figures10-14, study periods and size classifications for each indicator are also depicted ungrouped with standard error for each classification.

For both fish and macroinvertebrates, nearly 35% of stream miles were classified in poor condition over the 4-year study period. Also, for both indicators, the poor category increased to greater than 40% from 2008-2009 and decreased to less than 25% from 2010-2011. A notable difference between the indicators is the higher percentage of stream miles in fair condition as opposed to good condition. For all study periods, the percentages of stream miles in fair condition are greater than 40% for macroinvertebrates and less than 10% for fish. When considering stream size, a greater percentage of large river stream miles are in poor condition than small streams. For benthic macroinvertebrates, nearly 65% of large river miles are in poor condition, with approximately 20% in fair or good condition. Conversely, in small streams, approximately 25% of stream miles are poor or good condition, while nearly 50% are in fair condition. Likewise, for fish, nearly 50% of large river miles are in poor condition and nearly 35% in good condition. In small streams, greater than 75% of miles are in good condition, while approximately 30% are in poor condition.

A relative small percentage of miles are classified in poor condition for benthic algae. For the 4-year study period, approximately 10% of miles are in poor condition, with greater than 75% of miles in good condition. In 2008-2009, the percentage in poor condition decreases to less than 5%, while the percentage in good condition increases to nearly 85%. However, in 2010-2011, the percentage in poor condition nearly doubles to greater than 20%, with greater than 65% in good condition. As with fish and macroinvertebrates, a greater percentage of large rivers (22%) than small streams (6%) are in poor condition.

For phytoplankton, or sestonic algae, the percentage of streams in poor condition across study years varies from nearly 20% (2008-2009) to nearly 30% from 2010-2011. The percent in good condition is approximately 55% for all study periods. Conversely, stream size varies significantly for poor and good condition. Approximately 60% of large river miles are in poor condition as compared less than 10% of small river miles. Conversely, less than 20% of large river miles are in good condition for sestonic algae, while nearly 70% of small river miles are considered in good condition.

Figure 8. Stacked percentages of condition class estimates for study periods grouped by biological indicators.

Figure 9. Stacked percentages of condition class estimates for stream size grouped by biological indicators.

Figure 10. Biological indicator condition extent estimated statewide from 2008-2011. Upper and lower bounds represent a 95% confidence interval.

Figure 11. Biological indicator condition extent estimated statewide for larger streams and rivers (Strahler Order > 4) from 2008-2011. Upper and lower bounds represent a 95% confidence interval.

Figure 12. Biological indicator condition extent estimated statewide for smaller streams and rivers (Strahler Order < 5) from 2008-2011. Upper and lower bounds represent a 95% confidence interval.

Figure 13. Biological indicator condition extent estimated statewide from 2008-2009. Upper and lower bounds represent a 95% confidence interval.

Figure 14. Biological indicator condition extent estimated statewide from 2010-2011. Upper and lower bounds represent a 95% confidence interval.

Stressor Extent

Statewide condition extent estimates were made for total nitrogen, conductivity, turbidity, metal toxicity, sedimentation, and instream and vegetative cover. Estimates employed a variety of NRSA and Omernick level III ecoregion screening levels. For each stressor except metals toxicity, the condition was categorized as good, fair, or poor based on methodology described in the "Methods" section, and percentages for each condition category are based on "percent of total miles".

In Figures 15-16, good/fair/poor estimates for nutrients, conductivity and turbidity are grouped for each stressor by both study periods and size. In Figures 17-21, study periods and size classifications for each indicator are also depicted ungrouped with standard error for each classification. Phosphorus extent in poor condition is generally 30-40%, regardless of study period or source of screening limit, while the percent of total miles in good condition ranges from 40-50%. Generally, poor condition is lower for the NRSA screening limits, but is not significantly different. When considering stream size, large streams (approximately 75%) have a significantly higher percentage of miles in poor condition than small streams (10-25%). For total nitrogen, the difference between the sources of screening limits and study periods are more dramatic, but still not significantly different. For the NRSA screening limit, the percent of miles in poor condition ranges from less than 15% from 2008-2009, as opposed to nearly 25% from 2010-2011. For all study periods, good condition is greater than 50%. A similar pattern is evident with the ecoregion screening limit, with poor condition ranging from 25% (2008-2009) to nearly 40% from 2010-2011, and good condition ranging from nearly 50% to as low as approximately 25% during the same periods. Unlike total phosphorus, stream size is not significant when considering percent of miles in poor condition, with large ranging from 30-40% and small from 10-25%. However, the percent in good condition is significantly different with size. Large rivers range from 15-25%, with small streams ranging from 55-65% in good condition. Conductivity is generally not significantly different between study periods or sources of screening limits. Poor condition ranges from 10-20% from 2008-2009, and increases approximately 22% in the 2010-2011 study period. For the NRSA values, good condition is ranges from 60-65%, regardless of period. However, when using ecoregion screening limits, good condition during the 2010-2011 period shows a significant decrease to approximately 25%, while the 2008-2009 period is approximately 55%. As with nutrients, condition is significantly different when comparing streams to rivers. The percent of river miles in poor condition ranges from 40-55%, while streams are approximately 5% for both screening limits. Conversely, the percent of stream miles in good condition ranges from 55% to 80%, as opposed to 15-30% in large rivers. For turbidity, period is not significant, with poor condition ranging from 10- 30%, and good at 30-35% for both periods. However, the percent of river miles (37%) in poor condition is significantly different from small streams (9%). The percentage in good condition is much closer with nearly 30% in large rivers and 35% in small streams.

The extent of various habitat stressors is depicted in Figure 22. Instream and riparian vegetative cover are considered for only the 2008-2009 period, with no delineation between waterbody sizes. Poor condition ranges from 5% for riparian to 15% for instream cover. Good condition is 65-70% for both. Excess sedimentation is not significantly different when considering waterbody size. Poor condition ranges from greater than 25% in streams to 35% in rivers, with the percent in good condition at nearly 35% in both. Study period is significantly different, with poor condition ranging from 15% (2008-2009) to greater than 50% from 2010-2011. Good condition is not significantly different, but does range from less than 20% (2010-2011) to greater than 50% in 2008-2009 study period. The percent of miles in fair condition is ranges from 30-40%, regardless of study period.

Finally, the extent of metals toxicity is represented in Figure 23. Miles in poor condition generally ranges from 10-15%. While no stressors are significantly different, more miles appear to be affected by selenium than any other metal.

Figure 15. Stacked percentages of condition class estimates for study periods grouped by stressors. (Refer to Table 7 for stressor descriptions.)

Figure 16. Stacked percentages of condition class estimates for stream size grouped by stressors. (Refer to Table 7 for stressor descriptions.)

Figure 17. Stressor extent estimated statewide from 2008-2011. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

Figure 18. Stressor extent estimated statewide for larger streams and rivers (Strahler Order > 4) from 2008-2011. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

Figure 19. Stressor extent estimated statewide for smaller streams and rivers (Strahler Order < 5) from 2008-2011. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

Figure 20. Stressor extent estimated statewide from 2008-2009. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

Figure 21. Stressor extent estimated statewide from 2010-2011. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

Figure 22. Sedimentation and other habitat stressors estimated statewide from 2008-2011. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

Figure 23. Metal toxicity extent estimated statewide from 2010-2011. Upper and lower bounds represent a 95% confidence interval. (Refer to Table 7 for stressor descriptions.)

RESULTS—RELATIVE RISK

Relative Risk Methodology

The concept of using relative risk to develop a relationship between biological condition and stressor extent was developed initially for the USEPA's National Wadeable Streams Assessment (USEPA, 2006). Van Sickle et al. (2006) drew upon a practice commonly used in medical sciences to determine the relationship of a stressor (e.g., high cholesterol) to a medical condition (e.g., heart disease). The method calculates a ratio between the number of streams with poor biological condition/high stressor concentration and those with poor biological condition/low stressor concentration. If the ratio is above 1, it indicates that biological condition is likely affected by high stressor concentrations (i.e., concentrations above a preset level). As the ratio increases beyond 1, the relative risk of the stressor increases (Van Sickle, 2004).

The following analyses include a comparison of a variety of stressors to biological conditions for fish, macroinvertebrates, and algal biomass. For each stressor, relative risk is determined for study period and/or waterbody size. The analysis uses a binomial designation of good/poor for condition and high/low for stressor concentration. These binomial designations are then placed in a two-way contingency table to determine relative risk. Two initial ratios are determined. The ratio for poor condition given high stressor concentration is compared to the total number of sites having high stressor concentration, regardless of condition. Likewise, the ratio for poor condition given low stressor concentration is compared to the total number of sites having low stressor concentrations, regardless of condition. These two ratios are then used to calculate relative risk. For each indicator and stressor, the good and fair conditions were collapsed into a good condition for purposes of calculating relative risk. Significant relative risk will be determined by applying a 95% confidence, which must remain above 1.0 for risk to be considered significant.

Relative Risk to Fish Condition

The relative risks of various stressors to fish condition are represented in Figures 24-29. The relative risk of poor fish condition is generally greater than 1 when most stressors are in poor condition. However, few are not significant, regardless of study period or size. For the 2008-2009 study period, the ecoregion total nitrogen screening limit shows a significant relative risk of nearly 2.5 to fish condition (Figure 24). Likewise, if the NRSA conductivity is in poor condition, the risk of poor fish condition is 4.7 times greater during 2010-2011 period (Figure 26) and 2.9 times greater in small streams (Figure 27). Additionally, the risk for poor fish condition is 1.7 times greater in large rivers when turbidity is in poor condition (Figure 27). When excess sediment leads to poor condition, poor fish condition is 2.1 times more likely in large rivers and 2.3 times during the 2008- 2009 study period (Figure 28). Metals toxicity demonstrates no significant relative risk to fish condition (Figure 29).

Figure 24. Relative risk of nutrient stressors affecting poor fish condition by study period. (upper/lower bounds represent a 95% confidence interval-CI) (* = significant relative risk-RR)

Figure 25. Relative risk of nutrient stressors affecting poor fish condition by waterbody size. (upper/lower bounds represent a 95% CI) ($* =$ significant RR)

Figure 26. Relative risk of conductivity and turbidity stressors affecting poor fish condition by study period. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 27. Relative risk of conductivity and turbidity stressors affecting poor fish condition by waterbody size. (upper/lower bounds represent a 95% CI) ($*$ = significant RR)

Figure 28. Relative risk of sediment and habitat stressors affecting poor fish condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 29. Relative risk of metal toxicity stressors affecting poor fish condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Relative Risk to Macroinvertebrate Condition

The relative risks of various stressors to macroinvertebrate condition are shown in Figures 30-35. As with fish, the relative risk of poor macroinvertebrate condition is generally greater than 1 when most stressors are in poor condition, but unlike fish, many demonstrate significant risk. With the exception of the NRSA screening limit during study certain periods, the risk of poor macroinvertebrate condition is 2.3 to 3.5 times greater with poor total phosphorus condition and 1.9 to 4.3 times greater with poor total nitrogen condition (Figure 30). For stream size, small streams with poor total phosphorus condition are 3.4 to 5.8 times more likely to have poor macroinvertebrate condition (Figure 31). Poor total nitrogen condition, regardless of size, and total phosphorus in large rivers do not pose a significant relative risk to macroinvertebrate condition. When conductivity is in poor condition, all waterbodies are 2.4 to 3.2 times more likely to have poor macroinvertebrate condition, and from 2010-2011, poor condition was 2.9 times more likely when turbidity was in poor condition (Figure 32). Depending on the screening limit, risk of poor macroinvertebrate condition is 3.7 times greater in small streams and 1.7 times greater in large rivers when conductivity condition is poor (Figure 33). Turbidity demonstrates no significant relative risk to macroinvertebrate condition when considering waterbody size. The risk of poor macroinvertebrate condition is 2.3 times more likely when riparian vegetative cover is poor, and from 2008-2009, was 2.6 greater with excess sedimentation (Figure 34). Large rivers are also 1.5 times more likely to show poor condition with excess sedimentation. Finally, as with fish, metals toxicity demonstrates no significant relative risk to macroinvertebrate condition (Figure 29).

Figure 30. Relative risk of nutrient stressors affecting poor macroinvertebrate condition by study period. (upper/lower bounds represent a 95% CI) $(* =$ significant RR)

Figure 31. Relative risk of nutrient stressors affecting poor macroinvertebrate condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 32. Relative risk of conductivity and turbidity stressors affecting poor macroinvertebrate condition by study period. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 33. Relative risk of conductivity and turbidity stressors affecting poor macroinvertebrate condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 34. Relative risk of sediment and habitat stressors affecting poor macroinvertebrate condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 35. Relative risk of metal toxicity stressors affecting poor macroinvertebrate condition by waterbody size. (upper/lower bounds represent a 95% CI) ($* =$ significant RR)

Relative Risk to Benthic Algae Condition

The relative risks of various stressors to benthic algae condition are represented in Figures 36-41. Nutrients show very little significant relative risk, regardless of study period or waterbody size (Figures 36 and 37). For the 4-year study period, benthic algae condition was 3.3 times more likely to be poor when NRSA total nitrogen was in poor condition (Figure 36). Likewise, over the entire study period, poor conductivity condition was 3.0 to 4.5 times more likely to lead to excessive benthic algal growth (Figure 38). When the ecoregion conductivity was high, the likelihood of poor condition in the population increased by 10.2 times from 2008-2009 (Figure 38) and 9.6 times in small streams (Figure 39). Poor turbidity and habitat condition, as well as excess sedimentation, pose no significant relative risk to benthic algae condition (Figures 38-40). Interestingly, excess benthic algal growth is more likely when lead (3.9) and selenium (2.9) are above the applicable chronic toxicity criteria (Figure 41).

Figure 36. Relative risk of nutrient stressors affecting poor benthic algae condition by study period. (upper/lower bounds represent a 95% CI) ($* =$ significant RR)

Figure 37. Relative risk of nutrient stressors affecting poor benthic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 38. Relative risk of conductivity and turbidity stressors affecting poor benthic algae condition by study period. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 39. Relative risk of conductivity and turbidity stressors affecting poor benthic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 40. Relative risk of sediment and habitat stressors affecting poor benthic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 41. Relative risk of metal toxicity stressors affecting poor benthic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Relative Risk to Sestonic Algal Condition

The relative risks of various stressors to sestonic algae condition are represented in Figures 42-47. Regardless of study period, poor total phosphorus condition increases by 4.7-15.6 times the risk of poor sestonic algae condition (Figure 42). When the NRSA total nitrogen screening limit is in poor condition, the risk of poor condition increases by 2.8-3.4 times during the 4-year study period as well as the 2010-2011 period. In large rivers, the risk of poor sestonic algae condition increases by 2 to 3.7 times when total phosphorus is in poor condition and 1.4 times when total nitrogen is poor (Figure 43). Conversely, in small rivers, significant risk is confined to NRSA total phosphorus, with the risk of excess sestonic algal growth increased by 6.6 times. With poor conductivity condition, the risk of increased algal growth increased by 3 to 6.6 times, and poor turbidity condition increased risk by 2.7 to 4.5 times, during the 4-year and the 2008-2009 study periods (Figure 44). In small streams, high conductivity increased by 5.3 times the risk for excess sestonic algal growth (Figure 45), while large rivers showed no significant relative risk related to conductivity. There was not significant relative risk to poor turbidity condition in large or small waterbodies. Excess sediment and poor riparian vegetative cover did not significantly increase relative risk (Figure 46). However, poor instream cover increased the likelihood of excessive sestonic algal growth by 2.9 times. Finally, lead concentrations above the chronic criterion increased the likelihood of excessive algal growth by 3.2 times (Figure 47).

Figure 42. Relative risk of nutrient stressors affecting poor sestonic algae condition by study period. (upper/lower bounds represent a 95% CI) $(* =$ significant RR)

Figure 43. Relative risk of nutrient stressors affecting poor sestonic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) ($* =$ significant RR)

Figure 44. Relative risk of conductivity and turbidity stressors affecting poor sestonic algae condition by study period. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 45. Relative risk of conductivity and turbidity stressors affecting poor sestonic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 46. Relative risk of sediment and habitat stressors affecting poor sestonic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

Figure 47. Relative risk of metal toxicity stressors affecting poor sestonic algae condition by waterbody size. (upper/lower bounds represent a 95% CI) (* = significant RR)

DISCUSSION AND RECOMMENDATIONS

Oklahoma's Integrated Water Quality Report

Oklahoma's environmental agencies gather and assess data across the state for a wide variety of biological, chemical, and physical water quality indicators. One purpose of these data collections is to meet federal Clean Water Act requirements to compile a list of impaired waterbodies and determine the condition of all of these waters. These reports are compiled to the biannual Oklahoma Water Quality Assessment Integrated Report (ODEQ, 2010).

The current study benefits this report in several ways. First, this report marks Oklahoma's second and third statistically based assessments of the condition of Oklahoma's lotic waters. The OWRB recommends that this report be adopted into the 305(b) section of the 2012 or 2014 integrated report. Included graphics can be used to show overall statewide and regional condition. Second, individual lotic waterbodies not yet included in Oklahoma's Integrated Report (ODEQ, 2010) now have some level of assessment. The OWRB regularly submits waters for inclusion on Oklahoma's 303(d) list, and will do so again in October 2013. As a part of OWRB's submission, waterbodies assessed as part of this study will be included for consideration as not only category 5 (impaired), but as category 3 (not impaired for some uses). Because of assessment rules housed in Oklahoma's Continuing Planning Process (CPP) (ODEQ, 2012) and USAP (OWRB, 2012b), certain water quality parameters will not be included as part of the assessment. Most of Oklahoma's assessment protocols require that certain data requirements be met including the number of samples required to make an assessment determination. Protocols were developed to either assess short-term or long-term exposure. Short-term exposure protocols are written as percent exceedances, with typically a minimum of ten samples required. Long-term exposure protocols are based upon some measure of central tendency, but typically require a minimum number of samples to calculate the applicable descriptive statistic. Some exceptions to these rules include biological assessments, application of the sediment criteria, and a single sample maximum of 200 mg/m³ for benthic chlorophyll-a. All other parameters included in this study will not be included in assessments for the impaired waters list but will be made publicly available in the event that another entity can include the data in their assessment. To ensure inclusion of relevant data, stations will be placed in the most current version of the OWRB Assessment Workbook (OWRB 2013c), which is not only used to assess waters for the Oklahoma Integrated Report but for the OWRB's Beneficial Use Monitoring Program (OWRB, 2013a)

Differences in Indicator/Stressor Levels

The current study allows for unique analysis between both study periods and waterbody size. Differences in poor condition of both indicators and stressors are presented in Table 10. The analysis simply compares the differences in percent of total miles in poor condition, and establishes significant difference between periods or size if the 95% confidence interval does not overlap the calculated percentage of the other subcategory. For example, for fish, the confidence intervals of period percentages overlap but do not overlap the calculated percentage. Additionally, the arrows in the trend column merely indicate the direction of a potential trend.

For indicators, both fish and macroinvertebrates demonstrate a downward trend in poor condition between study periods, with only the fish having a significant downward trend. Conversely, both algal indicators show an upward trend, with only the benthic algae trend having significance. Likewise, all but one of the total phosphorus stressors shows an upward trend between the two study periods, with only turbidity and sediment having a significant trend. Notably, environmental conditions, particularly drought, became more acute in 2010-2011, and high water was an issue

during a portion of the 2009 index period. Otherwise, no other notable differences exist between the two periods, except the MMI used to analyze to macroinvertebrates, which could account for the difference in poor condition between the two periods. Lastly, when comparing large to small waterbodies, all indicators and stressors have a larger percentage of large river miles in poor condition that small river miles. And, with the exception of sediment, all differences are significant. Likely, this exists for several reasons. First, larger rivers and streams carry much heavier pollutant loads because they have a much larger area of input. Second, the development and refinement of reference condition, metrics, and stressor criteria/screening limits need continued development at both ecoregion and size scales. Data exists to perform these tasks and would eliminate much of the potential noise that is present in current assessments.

Table 10. The percentage of indicators and stressors in poor condition compared between study periods, as well as large and small waterbodies. Arrows show direction of potential trend (** = significant at alpha of 0.95)

Attributable Risk

To determine the actual affect a stressor has on a particular biological indicator, relative risk analyses were made for each stressor-indicator pair and presented in the results section of this report. However, is there a way to determine how much affect a proportional reduction in a stressor would have on the incidence of poor condition in an indicator? Attributable risk provides an elimination scenario to investigate this relationship and potential beneficial outcomes of reduction (Sickle and Paulsen, 2008). Although assailable assumptions are made about causality and the analysis requires elimination of the stressor, it is still a useful extension of the stressor extent and risk models already used in probability assessments. As reported in the draft NRSA report:

"Attributable risk represents the magnitude or importance of a potential stressor and can be used to help rank and set priorities for policymakers and managers. Attributable risk is derived by combining relative extent and relative risk into a single number for purposes of ranking. Conceptually, attributable risk provides an estimate of the proportion of poor biological conditions that could be reduced if high levels of a particular stressor were eliminated. This risk number is presented in terms of the percent of length that could be improved" (USEPA, 2013).

The results of attributable risk for the current Oklahoma studies are provided in Figures 49-52. In order to provide a meaningful analysis, an assumption was made that if relative risk was not significant, then calculating of an elimination scenario was not meaningful. Therefore, pollutant elimination analyses were only performed where stressor/indicator relative risk was significant. Confidence intervals were also calculated for each risk analysis, and significant potential reduction only exists where the upper confidence bound does not equal the original percent in poor condition. For example, in Figure 49, an elimination of turbidity could reduce poor condition for fish in large rivers by approximately 10%. However, upper confidence bound is not lower than the original percentage in poor condition, so the potential reduction is effectively not different from "0".

Notably, for fish, elimination of sediment in large rivers could create a significant reduction of poor condition in fish as could reduction in conductivity (Figure 49). For macroinvertebrates, elimination of both total phosphorus and total nitrogen could have a significant effect on poor condition (Figure 50). The elimination of phosphorus in small streams results in a nearly 14% lowering of the percent of miles in poor condition. As with fish, the elimination of conductivity is significant in some scenarios. Sestonic algal condition shows potential promise with a variety of pollutant elimination scenarios (Figure 52). Turbidity, conductivity, and nutrients all show some significant results. Of particular interest, many of the total phosphorus measures show significant potential reduction in sestonic algal growth. For example, in large rivers, the elimination of phosphorus would reduce the percent of river miles in poor condition by greater than 25 to 40%. There is no significant pollutant elimination scenarios related to benthic algae condition (Figure (51).

Interestingly, the elimination of conductivity is consistently significant in reducing the prevalence of poor indicator condition. Because of Oklahoma's significant conductivity gradient, this is to be expected. However, it is yet another indication of the need for refinement and further regionalization of reference condition and biological criteria, as well as the potential effect of dewatering and drought in alluvial systems. Likewise, the potential that the elimination of phosphorus would have on biological condition is prevalent throughout the analysis, regardless of study period, waterbody size, or screening limit source.

Future Plans

In terms of monitoring, probabilistic design has been completely integrated into both the OWRB and OCC monitoring programs (OWRB, 2012d). The OWRB is currently participating in the 2013-2014 National Rivers and Streams Assessment (NRSA) and will use data from it to provide an update to the current report. Also, the fourth two-year statewide study will begin in 2015 (OWRB, 2013b). Substantive changes to the program will include: 1) use of the NRSA protocols for large Wadeable and non-wadeable waterbodies, 2) use of NRSA habitat protocols for wadeable streams in concert with the current RBP habitat protocol, 3) inclusion of a second winter macroinvertebrate index period, 4) development of a periphyton taxonomic assemblage, 5) assessments at aggregated ecoregion scales used in the 2005-2007 assessment (OWRB, 2009), and 6) change/trend analyses through the use of revisit sites. Dependent upon future funding, additional plans are also in the works for future regionally based studies, similar to the Illinois River Basin Project (OWRB, 2010a).

Figure 48. Potential reduction to poor condition of fish based on the attributable risk of stressors having significant relative risk. (upper/lower bounds represent a 95% confidence interval-CI)

Figure 49. Potential reduction to poor condition of macroinvertebrates based on the attributable risk of stressors having significant relative risk. (upper/lower bounds represent a 95% confidence interval-CI)

Figure 50. Potential reduction to poor condition of benthic algae based on the attributable risk of stressors having significant relative risk. (upper/lower bounds represent a 95% confidence interval-CI)

Figure 51. Potential reduction to poor condition of sestonic algae based on the attributable risk of stressors having significant relative risk. (upper/lower bounds represent a 95% confidence interval-CI)

REFERENCES

Kincaid, Thomas. 2013. Personal Communication of Adjusted Site Weights. USEPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division, Corvallis, OR.

Minitab, Inc. 2013. Minitab Statistical Software version 16.2.4. State College, PA.

Mitchell, R. 2013. Personal Communication of 2008-2009 NRSA Chlorophyll Data. USEPA, Office of Water, Office of Wetlands, Oceans, and Watersheds (OWOW), Assessment and Watershed Protection Division, Washington, DC.

North American Commission for Environmental Cooperation. 1997. *Ecoregions of North America (contains Omernik Level I, II, and III)*. GIS map files. North American Commission for Environmental Cooperation, Montreal, Quebec, Canada.

Oklahoma Conservation Commission. 2005a. *Analysis of Oklahoma Conservation Commission Physiochemical and Biological Data Toward Determination of High Quality Sites*". Oklahoma City, OK.

Oklahoma Conservation Commission. 2005b. *Small Watershed Rotating Basin Monitoring Program Year 1: Neosho-Grand and Upper Canadian Basins*. Oklahoma City, OK.

Oklahoma Conservation Commission. 2006a. *Small Watershed Rotating Basin Monitoring Program Year 2: Upper North Canadian, Cimarron, and Upper Arkansas Basins*. Oklahoma City, OK.

Oklahoma Conservation Commission. 2006b. *Small Watershed Rotating Basin Monitoring Program Year 3: Lower Arkansas, Lower Canadian, and Lower North Canadian Basins*. Oklahoma City, OK.

Oklahoma Conservation Commission. 2007. *Small Watershed Rotating Basin Monitoring Program Year 4: Washita and Upper Red Basins*. Oklahoma City, OK.

Oklahoma Conservation Commission. 2008. *Small Watershed Rotating Basin Monitoring Program Year 5: Lower Red Basin*. Oklahoma City, OK.

Oklahoma Conservation Commission. 2009. Personal Communication. Oklahoma City, OK.

Oklahoma Department of Environmental Quality. 2007. *Quality Management Plan*. QMP Track No. 00-182. Oklahoma City, OK.

Oklahoma Department of Environmental Quality. 2010. *The State of Oklahoma 2010 Water Quality Assessment Integrated Report*. Oklahoma City, OK.

Oklahoma Department of Environmental Quality. 2012. *Continuing Planning Process*. Oklahoma City, OK.

Oklahoma Water Resources Board. 1999. *Technical Report 99-3: Standard Operating Procedures for Stream Assessments and Biological Collections Related to Biological Criteria and Development*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2004. *Standard Operating Procedure for the Use of Floats to Determine Stream Discharge.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2005. *Standard Operating Procedure for the Measurement of Stream Discharge.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2006a. *Development of a Stream/River Probabilistic Sampling Network for the State of Oklahoma: FY-2003 Section 104(b)3 Project 2, USEPA Region 6.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2006b. *Standard Operating Procedure for the Collection, Filtration, and Extraction of Benthic and Sestonic Chlorophyll-a Samples in Streams.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2009. *Implementation of a Stream/River Probabilistic Monitoring Network for the State of Oklahoma: FY-2005 Section 104(b)3 Regional Environmental Monitoring and Assessment Program Study*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2010a. *Probabilistic Monitoring in the Illinois River Watershed to Determine Multi-Assemblage Biotic Condition and Stressor Relationships*: *FY-2005 Section 104(b)3 Water Quality Cooperative Agreement*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2010b. *Standard Operating Procedure for the Collection of Fish in Streams.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2010c. *Standard Operating Procedure for the Collection of Habitat in Streams.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2010d. *Standard Operating Procedure for the Collection of Macroinvertebrates in Streams.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2010e. *Standard Operating Procedure for the Collection of Water Quality Samples in Streams*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2011. *Water Information Management System.* Oklahoma City, OK.

Oklahoma Water Resources Board. 2012a. *Fish and Macroinvertebrate Assessment Workbook v.3*. Oklahoma City, OK.

Oklahoma Water Resource Board. 2012b. *Implementation of Oklahoma's Water Quality Standards, Chapter 46, Subchapter 15: Use Support Assessment Protocols (USAP).* OAC 785:46-15. Oklahoma City, OK.

Oklahoma Water Resources Board. 2012c. *Oklahoma Water Quality Standards*. OAC 785:45. Oklahoma City, OK.

Oklahoma Water Resources Board. 2012d. *Oklahoma's Water Quality Monitoring Strategy*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2013a. *Oklahoma's Beneficial Use Monitoring Program Annual Report*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2013b. *Statewide Surface Water Probabilistic Monitoring Network Quality Assurance Project Plan: FY-2013Section 106 Water Pollution Control Program*. Oklahoma City, OK.

Oklahoma Water Resources Board. 2013c. *Water Quality Assessment Workbook v.2*. OKC, OK.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers 77*:118–125.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, R. M. Hughes. 1989. *Rapid Bioassessment Protocols for Use in Streams and Rivers*. USEPA/444/4-89-001. U. S. E.P.A, Assessment and Watershed Protection Division, Washington, D. C.

R Foundation for Statistical Computing. 2013. *R version 3.0.0 -- "Masked Marvel".* Platform: i386-w64-mingw32/i386 (32-bit)

Stevens, Jr., D.L. 1997. Variable density grid-based sampling designs for continuous spatial Populations. *Environmetrics.* 8:167-195.

Stevens, Jr., D. L., and A. R. Olsen. 2004. Spatially-balanced sampling of natural resources. Journal of American Statistical Association 99(465): 262-278.

USEPA. 2000a. *Information Supporting the Development of State and Tribal Nutrient Criteria for Rivers and Streams in Nutrient Ecoregion IX: Southeastern Temperate Forested Plains and Hills*. USEPA. Office of Water, Office of Science and Technology, Health and Ecological Criteria Division, Washington, D.C.

USEPA. 2000b. *Information Supporting the Development of State and Tribal Nutrient Criteria for Rivers and Streams in Nutrient Ecoregion XI: Central and Eastern Forested Uplands*. Office of Water, Office of Science and Technology, Health and Ecological Criteria Division, Washington, D.C.

USEPA. 2001a. *Information Supporting the Development of State and Tribal Nutrient Criteria for Rivers and Streams in Nutrient Ecoregion IV: Great Plains Grass and Shrublands*. Office of Water, Office of Science and Technology, Health and Ecological Criteria Division, Washington, D.C.

USEPA. 2001b. *Information Supporting the Development of State and Tribal Nutrient Criteria for Rivers and Streams in Nutrient Ecoregion V: South Central Cultivated Great Plains*. Office of Water, Office of Science and Technology, Health and Ecological Criteria Division, Washington, D.C.

USEPA. 2005. *The Recommended Elements of a State Monitoring Program*. <http://www.epa.gov/owow/monitoring/elements/elements.html> Office of Water, Office of Wetlands, Oceans, and Watersheds, Washington, D.C.

USEPA. 2006. *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams*. USEPA 841-B-06-002. Office of Research and Development, Office of Water, Office of Science and Technology, Health and Ecological Criteria Division, Washington, D.C.
USEPA. 2009. National Rivers and Streams Assessment: Field Operations Manual. EPA 841/B-04/004, Office of Water and Office of Environmental Information, US Environmental Protection Agency, Washington, DC.

USEPA. 2012. USEPA Natioanal Aquatic Resource Surveys Sharefile Site: [https://nars.sharefile.com/.](https://nars.sharefile.com/)

USEPA. 2013. National Rivers and Streams Assessment 2008-2009: A Collaborate Survey. EPA 841/D-13/001, Office of Water and Office of Environmental Information, US Environmental Protection Agency, OWOW, Washington, DC.

Van Sickle, J. 2004. Regional Assessment of Stressor Severity: A Relative Risk Approach. National Health and Environmental Effects Research Laboratory, Western Ecology Division, Corvallis, OR. Presented at the NW Biological Assessment Workgroup 15th Annual Meeting.

Van Sickle, J., J.L. Stoddard, S.G. Paulsen, and A.R. Olsen. 2006. Using relative risk to compare the effects of aquatic stressors at a regional scale. *Environmental Management 38*(6):1020–1030.

Van Sickle, J., and S.G. Paulsen. 2008. Assessing the Attributable Risks, Relative Risks, and Regional Extents of Aquatic Stressors. *Journal of the North American Benthological Society. 27*(4):920–931.

Van Sickle, J. 2012. *R Software Workshop Scripts 1-6 from Environmental Statistics for Enviroscience, Inc.; USEPA/NARS Contract PR-R6-12-00504.* R Statistical Workshop, USEPA Region 6, February 2013. Dallas, TX.

Woods, A.J., Omernik, J.M., Butler, D.R., Ford, J.G., Henley, J.E., Hoagland, B.W., Arndt, D.S., and Moran, B.C. 2005. Ecoregions of Oklahoma (color poster with map, descriptive text, summary tables, and photographs): Reston, Virginia, U.S. Geological Survey (map scale 1:1,250,000).

APPENDIX A–TARGET STATION METADATA

Table 11. Appendix A—Metadata for Target Sites.

APPENDIX B – CONDITION CLASSES

Table 12. Appendix B—Biological Indicator Condition Classes.

Table 13. Appendix B—Stressor Indicator Condition Classes.

APPENDIX C – DATA

Table 14. Appendix C—Fish Assessment Information.

Site ID	Sample_Type	Habitat	Sp_Rich	EPT Rich	%EPT	%DOM	$S-D$	HBI	%REF	Classification	
FW08OK053	RBP	SSV	16.00	3.00	13.11	44.22	$\overline{3}$	6.67	58.70	Slightly Impaired	
FW08OK067	LR	THAB	13.00	5.00	62.35	45.29	3	5.50	110.53	Reference	
FW08OK067	LR	THAB	11.00	5.00	58.79	58.79	3	5.12	102.63	Reference	
FW08OK067	LR	THAB	10.00	1.00	14.04	71.05	$\overline{2}$	7.08	55.26	Slightly Impaired	
FW08OK067	LR	THAB	10.00	2.00	42.55	52.48	3	5.76	78.95	Slightly Impaired	
FW08OK067	LR	SUB	18.00	9.00	46.83	39.68	4	5.05	126.32	Reference	
FW08OK067	LR	SUB	8.00	0.00	0.00	70.83	$\overline{2}$	6.87	39.47	Moderately Impaired	
FW08OK067	LR	SUB	10.00	1.00	14.04	71.05	$\overline{2}$	7.08	55.26	Slightly Impaired	
FW08OK067	LR	SUB	12.00	1.00	2.67	49.33	$\overline{3}$	5.81	63.16	Slightly Impaired	
FW08OK068	RBP	WOOD	15.00	4.00	15.63	28.13	4	5.29	85.25	Non-Impaired	
FW08OK068	RBP	WOOD	18.00	3.00	10.19	47.22	3	7.32	88.89	Non-Impaired	
FW08OK070	LR	THAB	13.00	1.00	0.81	82.93	$\overline{2}$	8.33	56.76	Slightly Impaired	
FW08OK070	LR	SUB	21.00	0.00	0.00	53.77	4	8.13	72.97	Slightly Impaired	
FW08OK070	LR	SUB	14.00	1.00	3.21	57.69	3	7.75	64.86	Slightly Impaired	
FW08OK070	LR	THAB	14.00	1.00	3.21	57.69	$\overline{3}$	7.75	64.86	Slightly Impaired	
FW08OK070	LR	COMP	13.00	1.00	1.39	66.67	$\overline{2}$	7.49	56.76	Slightly Impaired	
FW08OK070	LR	COMP	15.00	4.00	15.63	28.13	$\overline{4}$	5.29	113.51	Reference	
FW08OK073	LR	THAB	14.00	9.00	65.65	44.27	$\overline{3}$	5.14	124.44	Reference	
FW08OK073	LR	SUB	11.00	2.00	9.09	80.17	$\overline{2}$	6.15	62.22	Slightly Impaired	
FW08OK073	LR	SUB	11.00	3.00	4.49	73.03	$\overline{2}$	6.28	71.11	Slightly Impaired	
FW08OK073	LR	SUB	16.00	10.00	78.79	61.62	$\overline{3}$	4.53	115.56	Reference	
FW08OK075	LR	SUB	11.00	2.00	8.33	50.00	3	7.28	51.61	Moderately Impaired	
FW08OK075	LR	THAB	17.00	2.00	1.10	85.71	$\overline{2}$	7.21	45.16	Moderately Impaired	
FW08OK075	LR	THAB	10.00	0.00	0.00	47.42	$\overline{3}$	5.04	45.16	Moderately Impaired	
FW08OK075	LR	THAB	21.00	4.00	5.56	34.26	4	6.82	64.52	Slightly Impaired	
FW08OK076	RBP	WOOD/SSV	6.00	0.00	0.00	95.80	$\mathbf{1}$	6.00	26.23	Moderately Impaired	
FW08OK076	RBP	WOOD/SSV	6.00	0.00	0.00	83.04	$\overline{2}$	7.83	44.44	Moderately Impaired	

Table 15. Appendix C—Macroinvertebrate Assessment Information (2010-2011 Samples).

Table 16. Appendix C—Habitat Assessment Information (2010-2011 Samples).

Source	Station ID	Sample Date	N, Total (mg/L)	P, Total (mg/L)	SpC (uS/cm2)	Turbidity (NTU)	Ses_Chla (mg/m3)	Ben Chla (mg/m2)	Cd, Dis (ug/L)	Cu, Dis (ug/L)	Pb, Dis (ug/L)	Se, TR (ug/L)	Zn, Dis (ug/L)
NRSA	FW08OK002	7/8/2008	0.460	0.035	475.9	25.1	2.464	9.7	ND	ND	ND	ND	ND
NRSA	FW08OK003	5/26/2009	1.406	0.049	2315.2	16.7	3.4133	28.7	ND	ND	ND	ND	ND
NRSA	FW08OK005	6/15/2009	1.061	0.196	9944.2	11.2	39.888	23.1	ND	ND	ND	ND	ND
NRSA	FW08OK006	7/21/2009	1.069	0.081	234.4	5.1	13.7943	65	ND	ND	ND	ND	ND
NRSA	FW08OK007	7/7/2008	0.663	0.068	165.9	36.6	6.6	8.4	ND	ND	ND	ND	ND
NRSA	FW08OK009	6/9/2009	1.312	0.260	1517.2	22.4	9.1	9.6	ND	ND	ND	ND	ND
NRSA	FW08OK010	6/16/2008	0.516	0.050	834.5	16.6	2.176	4.4	ND	ND	ND	ND	ND
NRSA	FW08OK011	7/15/2008	1.058	0.155	1867.0	45.0	55.44	40.3	ND	ND	ND	ND	ND
NRSA	FW08OK012	9/30/2008	0.413	0.053	107.9	18.8	2.1333	13.5	ND	ND	ND	ND	ND
NRSA	FW08OK013	6/16/2009	2.111	0.769	1568.1	21.7	17.84	41.3	ND	ND	ND	ND	ND
NRSA	FW08OK014	8/18/2008	1.489	0.348	2850.0	56.9	119.8983	190.6	ND	ND	ND	ND	ND
NRSA	FW08OK017	6/22/2009	0.222	0.005	2620.6	0.7	0.516	16.7	ND	ND	ND	ND	ND
NRSA	FW08OK018	9/22/2009	1.903	0.286	505.3	66.3	17.8252	16	ND	ND	ND	ND	ND
NRSA	FW08OK019	7/29/2008	0.634	0.092	3922.6	46.0	51.982	14	ND	ND	ND	ND	ND
NRSA	FW08OK020	9/22/2008	1.273	0.413	1033.5	9.7	1.576	93.5	ND	ND	ND	ND	ND
NRSA	FW08OK022	7/8/2009	1.438	0.344	920.6	30.7	46.84	10.5	ND	ND	ND	ND	ND
NRSA	FW08OK023	9/18/2008	0.311	0.027	828.9	17.9	3.3	10.2	ND	ND	ND	ND	ND
NRSA	FW08OK024	6/8/2009	0.814	0.061	2068.3	4.8	1.04	24	ND	ND	ND	ND	ND
NRSA	FW08OK025	9/21/2009	0.608	0.080	11247.0	11.1	32.44	32.7	ND	ND	ND	ND	ND
NRSA	FW08OK026	9/2/2009	0.365	0.063	47.5	18.1	9.8489	16.4	ND	ND	ND	ND	ND
NRSA	FW08OK027	8/17/2009	1.146	0.229	6616.6	76.4	58.4	20.5	ND	ND	ND	ND	ND
NRSA	FW08OK028	6/16/2009	2.173	0.603	523.9	34.0	8.3667	23	ND	ND	ND	ND	ND
NRSA	FW08OK030	7/6/2009	1.983	0.235	271.1	31.8	17.55	40.2	ND	ND	ND	ND	ND
NRSA	FW08OK031	5/18/2009	0.368	0.018	501.9	1.4	1.648	22.1	ND	ND	ND	ND	ND
NRSA	FW08OK032	7/20/2009	1.184	0.147	1113.0	40.1	90.24	18.3	ND	ND	ND	ND	ND

Table 17. Appendix C—Chemistry, Chlorophyll, and Metals Data.

APPENDIX D – USEPANATIONAL RIVERS AND STREAMS ASSESSMENT TECHNICAL DOCUMENTATION (USEPA, 2012)

The documents provided in Appendices D1-D5 are provided by the USEPA at the NARS sharefile site (USEPA, 2012). Because NARS/NRSA technical evaluations were used to process and assess much of the data included in this report, these technical documents are provided for information and reference purposes only. However, they should not be considered or used as the final version or cited as seen in this report. To obtain the most recent version of the USEPA NARS technical reports, contact the USEPA, Office of Water.

APPENDIX D1 – NRSASURVEY DESIGN

Background Information:

The design requirements for the National Rivers and Streams Assessment are to produce:

- Estimates of the 2008-2009 status of flowing waters nationally and regionally (9 aggregated Omernik ecoregions),
- Estimates of the 2008-2009 status of wadeable streams and non-wadeable rivers nationally and regionally (9 aggregated Omernik ecoregions),
- Estimates of the 2008-2009 status or urban flowing waters nationally,
- Estimates of the change in status in wadeable streams between 2008-2009 and 2004, nationally and regionally (9 aggregated Omernik ecoregions).

A secondary objective is to have each state sample approximately an equal number of sites (37- 38).

Target population:

The target populations consists of all streams and rivers within the 48 contiguous states that have flowing water during the study index period excluding portions of tidal rivers up to head of salt. The study index period extends from April/May to September and is generally characterized by low flow conditions. The target population includes the Great Rivers. Runof-the-river ponds and pools are included while reservoirs are excluded.

Sample Frame:

The sample frame was derived from the National Hydrography Dataset (NHD), in particular NHD-Plus. Attributes from NHD-Plus and additional attributes added to the sample frame that are used in the survey design include: (1) state, (2) EPA Region, (3) NAWQA Mega Region, (4) Omernik Ecoregion Level 3 (NACEC version), (4) WSA aggregated ecoregions (nine and three regions), (5) Strahler order, (6) Strahler order categories (1st, 2nd, …, 7th, and $8th$ +), (6) FCode, (7) Urban, and (8) Frame07.

The version of NHD-Plus used includes two separate Strahler order calculations, one that is included on the publicly available NHD-Plus version. The other Strahler order calculation (SO attribute name) more accurately reflects the true Strahler order and is used for the survey design. The StrahCat attribute collapses $8th$, $9th$, and $10th$ order rivers in to a single category.

The Urban attribute was created by intersecting a modified version of the Census Bureau national urban boundary GIS coverage with NHD-Plus. The Census Bureau's boundaries were buffered 100 meters to include a majority of stream features intersecting and coincident with urban areas. Where this buffer did not completely gather all the river features within the urban areas (rivers intersecting cities are excluded from the Census Bureau's urban areas), the NHD-Plus river area (polygon) features were clipped at a three kilometer buffer around the urban areas and combined with the buffered urban area to create the modified urban database. If a stream or river segment was within this boundary, it is designated as "Urban"; otherwise as "NonUrban".

FCODE is directly from NHD-Plus and is used to identify which segments in NHD were included in the sample frame. The attribute Frame07 identifies each segment as either "Include" or "Exclude". Frame07 was created so that segments included in the sample frame could be easily identified. FCODE values included in the GIS shapefile:

Included in FW08 sample frame (Frame07='Include'):

- 33400 Connector
- 33600 Canal/Ditch
- 42801 Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = At or Near
- 46000 Stream/River
- 46003 Stream/River (Intermittent)
- 46006 Stream/River (Perennial)

58000 Artificial Path (removed from dataset if coded through Lake/Pond and Reservoirs) Excluded in FW08 sample frame (Frame07='Exclude')

- 42800 Pipeline
-

42802 Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Elevated

- 42803 Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Underground
- 42804 Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Underwater
- 42806 Pipeline: Pipeline Type = General Case; Relationship to Surface = Elevated
- 4280 Pipeline: Pipeline Type = General Case; Relationship to Surface = Underground
- 42809 Pipeline: Pipeline Type = Penstock; Relationship to Surface = At or Near
- 42811 Pipeline: Pipeline Type = Penstock; Relationship to Surface = Underground
- 42813 Pipeline: Pipeline Type = Siphon
- 56600 Coastline

Rivers that had Strahler order greater than or equal to $5th$ order and had FCODE equal to 46003 (intermittent) were included in the FW08 sample frame for all states west of 96 degrees longitude (North Dakota to Texas and states west). This was done to ensure that all large rivers in the more arid west were included regardless of NHD-Plus intermittent code.

Survey Design:

The survey design consists of two major components in order to address the dual objectives of (1) estimating current status for all flowing waters and (2) estimating change in status for wadeable streams from the 2004 Wadeable Stream Assessment. These two components are termed: (1) NRSA design and (2) WSA_Revisit design. A Generalized Random Tessellation Stratified (GRTS) survey design for a linear resource is used for the NRSA design and a GRTS survey design for a finite resource is used for the WSA_Revisit design. The design includes reverse hierarchical ordering of the selected sites.

Stratification:

The survey design is explicitly stratified by state for the NRSA design. The original WSA design had several strata (EMAP West, New England, Virginia, Iowa, and remaining eastern states combined). The WSA_Revisit design ignores these strata in the selection of the subset of sites from the WSA to be revisited as part of the current NRSA design.

Multi-density categories:

A complex unequal probability selection process was used in each of the two components of the survey design. They are described separately.

NRSA Design:

Unequal probability categories are defined separately for wadeable streams ($1st$ to $4th$ order) and non-wadeable rivers (5th to 10th order). Note wadeable and non-wadeable are used to designate Strahler order classes and do not imply that the streams will actually be wadeable or non-wadeable. The expected sample size is 450 for wadeable streams and 900 for non-wadeable rivers.

For wadeable stream category, within each state unequal selection probabilities were defined for 1st, 2nd, 3rd, and 4th order streams so that an equal number of sites would occur for each order. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category.

For non-wadeable river category, unequal selection probabilities were defined for $5th$, $6th$, $7th$, and $8th$ + order rivers so that the expected number of sites would be 350, 275, 175, and 100 sites, respectively. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category.

Given these initial selection probabilities, the expected number of urban and non-urban sites was calculated to determine if at least 150 urban sites would be selected. Over 150 urban sites were expected so no additional adjustment was required to satisfy the urban design equirement.

The final adjustment of the selection probabilities was to adjust them to minimize the range in the number of sites across the 48 states while still meeting the other design requirements. Given a total of 1350 sites for the NRSA design, each state would sample 28 sites. This could not be achieved, although the range was able to be decreased.

WSA_Revisit Design:

The Wadeable Stream Assessment sampled 1390 sites between 2000 and 2004. To estimate change, 450 of these sites will be revisited as part of the 2008-9 Rivers and Streams assessment. The revisit design selects the 450 sites using unequal selection probabilities. Initially, all sites were assigned an equal selection probability of 1.

First, four intensification study regions were sampled as part of the WSA. These regions are the Wenatchee Watershed in Washington, Lower John Day and Deschutes watersheds in Oregon, Northern California coastal watersheds, and southern California coastal give the expected number of sites within a study region if a state-wide survey design was done without ntensification.

Second, the density of sites sampled for the EMAP-West portion of the WSA was greater than for the 36 eastern states. The selection probabilities were reduced for EMAP-West states to adjust for this. The density of sites in the Southern Appalachian aggregated ecoregion was less than in other eastern aggregated ecoregions as a result of the site

replacement process used in the WSA. The selection probabilities were increased for these sites as well. The latter also ensured that the final weights for these sites were not extreme.

Third, the selection probabilities developed above were then adjusted to achieve approximately an equal number of sites across all nine WSA aggregated ecoregions.

Fourth, the overall weight, inverse of selection probability, was calculated by multiplying the original WSA weight by the inverse of the above selection probability. This accounts for the fact that the WSA Revisit design is a two-stage sample of wadeable streams.

WSA_Revisit design weights and NRSA design weights associated with wadeable streams will have to be adjusted to account for fact that they are two independent survey designs of wadeable streams for the 48 states. This will be done after the sites are evaluated and sampled.

State Designs:

For states that have a current, compatible state-wide probability design that cover all flowing waters, an option is provided to use their sites instead of the flowing water design sites. Whether the option is exercised for a state, requires that (1) their state design be a probability survey design, (2) their target population of streams and rivers includes the target population for the NRSA, (3) their sample frame includes the NRSA sample frame, and (4) their design is implemented state-wide in 2008-2009. The state must also agree to measure all the indicators included in the National Rivers and Streams Assessment using the national field and laboratory protocols.

Oversample:

No over sample sites were selected for the WSA_Revisit design. The expectation is that all, or almost all, of the 450 sites selected will be sampled given they were sampled previously. For the NRSA design, the over sample is nine times the expected sample size within each state. The large over sample size was done to accommodate those states who may want to increase the number of sites sampled within their state for a state-level design.

Site Use:

Each stream/river selected to be sampled is given unique site identification (siteID) that consists of two parts: (1) NFW08 that identifies the sites as part of the 2008-9 National Rivers and Streams Assessment and (2) the two-letter state FIPS code followed by a number between 001 and 999 within each state. It critical this siteID be used in its entirety to make sure that the stream and rivers sites are correctly identified.

Sites are organized to be used within each state. If a stream or river site is evaluated and determined that it can not be sampled, then it is to replaced by another site within the state. Sites that are coded as 1^{st} , 2^{nd} , 3^{rd} and 4^{th} are to be replaced by over sample sites that are coded 1st, 2^{nd} , 3^{rd} or 4^{th} , ignoring order within this range. For example, a 2^{nd} order would be replaced by either a 1st, 2nd, 3rd or 4th order stream. Sites that are coded as 5th, 6th, 7th, 8th, 9^{th} , or 10th order are to be replaced by over sample sites that are coded 5th, 6th, 7th, 8th, 9th, or $10th$ order, ignoring order within this range. For example, a $5th$ order river would be replaced by a $5th$, $6th$, $7th$, $8th$, $9th$, or 10th order river. In each case the next lowest sitelD that is within the Strahler order set is used for the replacement.

Evaluation Process

The survey design weights that are given in the design file assume that the survey design is implemented as designed. Typically, users prefer to replace sites that can not be sampled with other sites to achieve the sample size planned. The site replacement process is described above. When sites are replaced, the survey design weights are no longer correct and must be adjusted. The weight adjustment requires knowing what happened to each site in the base design and the over sample sites. EvalStatus is initially set to "NotEval" to indicate that the site has yet to be evaluated for sampling. When a site is evaluated for sampling, then the EvalStatus for the site must be changed. Recommended codes are:

Statistical Analysis

Any statistical analysis of data must incorporate information about the monitoring survey design. In particular, when estimates of characteristics for the entire target population are computed, the statistical analysis must account for any stratification or unequal probability selection in the design. Procedures for doing this are available from the Aquatic Resource Monitoring web page given in the bibliography. A statistical analysis library of functions is available from the web page to do common population estimates in the statistical software environment R.

Literature Cited

Diaz-Ramos, S., D. L. Stevens, Jr, and A. R. Olsen. 1996. EMAP Statistical Methods Manual. EPA/620/R-96/002, U.S. Environmental Protection Agency, Office of Research and Development, NHEERL-Western Ecology Division, Corvallis, Oregon.

Stevens, D.L., Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics, 8:167-95.

Stevens, D.L., Jr. and Olsen, A.R. 1999. Spatially restricted surveys over time for aquatic resources. Journal of Agricultural, Biological, and Environmental Statistics, 4:415-428

Stevens, D. L., Jr., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of environmental resources. Environmetrics **14**:593-610.

Stevens, D. L., Jr., and A. R. Olsen. 2004. Spatially-balanced sampling of natural resources in the presence of frame imperfections. Journal of American Statistical Association:99:262-278.

Horn, C.R. and Grayman, W.M. (1993) Water-quality modeling with EPA reach file system. Journal of Water Resources Planning and Management, 119, 262-74.

Strahler, A.N. 1957. Quantitative Analysis of Watershed Geomorphology. Trans. Am. Geophys. Un. 38,913-920.

APPENDIX D2 – NRSAPHYSICAL HABITAT

Background Information

An assessment of river and stream (fluvial) physical habitat condition is a major component of the National Rivers and Streams Assessment (NRSA). Of many possible general and specific fluvial habitat indicators measured in the NRSA surveys in 2008-2009, the assessment team chose streambed stability & excess fine sediments, instream habitat cover complexity, riparian vegetation, and riparian human disturbances for its assessment. These four indicators are generally important throughout the U.S. Furthermore, the project team had reasonable confidence in factoring out natural variability to determine expected values and the degree of anthropogenic alteration of the habitat attributes represented by these indicators.

In the broadest sense, fluvial habitat includes all physical, chemical, and biological attributes that influence or sustain organisms within streams or rivers. We use the term *physical habitat* to refer to the structural attributes of habitat. NRSA made field measurements aimed at quantifying eight general attributes of physical habitat condition, including direct measures of human disturbance.

- Habitat Volume/Stream Size
- Habitat Complexity and Cover for Aquatic Biota
- Streambed Particle Size
- Bed Stability and Hydraulic Conditions
- Channel-Riparian and Floodplain Interaction
- Hydrologic Regime
- Riparian Vegetation Cover and Structure
- Riparian Disturbance

These attributes were previously identified during EPA's 1992 national stream monitoring workshop (Kaufmann 1993) as those essential for evaluating physical habitat in regional monitoring and assessments. They are typically incorporated in some fashion in regional habitat survey protocols (Platts et al. 1983, Fitzpatrick et al. 1998, Lazorchak et al. 1998, Peck et al. 2006, Peck et al., in press USEPA, 2004) and were applied in the previous National Wadeable Streams Assessment (WSA) and the Western Rivers and Streams Pilot (EMAP-W) surveys conducted between 2000 and 2005 (USEPA 2006, Stoddard et al. 2005a,b). The major habitat metrics used in those past assessments and considered in NRSA are listed and defined in Table 1. Some measures of these attributes are useful measures of habitat condition in their own right (e.g., channel incision as a measure of channel-riparian interaction); others are important controls on ecological processes and biota (bed substrate size), still others are important in the computation of more complex habitat condition metrics (e.g., bankfull depth is used to calculate Relative Bed Stability [*RBS*]). Like biological characteristics, most habitat attributes vary according to their geomorphic and ecological setting. Even direct measures of riparian human activities and disturbances are strongly influenced by their geomorphic setting. And even within a region, differences in precipitation, stream drainage area channel gradient (slope) lead to variation in many aspects of stream habitat, because those factors influence discharge, flood stage, stream power (the product of discharge times gradient), and bed shear stress (proportional to the product of depth and slope). However, all eight of the major habitat attributes can be directly or indirectly altered by anthropogenic activities.

NRSA follows the precedent of EMAP-W and WSA in reporting the condition of fluvial physical habitat condition on the basis of four habitat indicators that are important nationwide, can be reliably and economically measured, and their reference condition under minimal anthropogenic disturbance can be interpreted with reasonable confidence. These are: relative bed stability (RBS) as an indicator of bed sedimentation or hydrologic alteration, the areal cover and variety of fish concealment features as a measure of in-stream habitat complexity, riparian vegetation cover and structure as an indicator of riparian vegetation condition, and a proximity-weighted tally of streamside human activities as an indicator of riparian human disturbances (Paulsen et al., 2008).

In this document, we describe the approach taken by NRSA for assessing physical habitat condition in rivers and streams based on the four above-mentioned indicators. We also examine the rationale, importance, and measurement precision of each of these indicators, including the analytical approach for estimating reference conditions for each. Reference conditions for each indicator were interpreted as their expected value in sites having the least amount of anthropogenic disturbance within appropriately stratified regions. In most cases, we also refine the expected values as a function of geoclimatic controlling factors within regions. Finally, we examine patterns of association between physical habitat indicators and anthropogenic disturbance by contrasting habitat indicator values in least- moderate- and most-disturbed sites nationally and within regions.

Methods

2.1 Physical Habitat Sampling and Data Processing

In the wadeable streams sampled in NRSA, field crews took measurements while wading the length of each sample reach (Peck et al. 2006); in non-wadeable rivers, these measurements were made from boats (Peck et al., in press). Physical habitat data were collected from longitudinal profiles and from 11 cross-sectional transects and streamside riparian plots evenly spaced along each sampled stream reach (U.S. EPA 2007). The length of each sampling reach was defined proportional to the wetted channel width and measurements were placed systematically along that length to represent the entire reach. Sample reach lengths were 40 times the wetted channel-width (ChW) long in wadeable streams, with a minimum reach length of 150 m for channels less than 3.5 m wide. In non-wadeable rivers, reach lengths were also set to 40 ChW with a maximum length of 2,000 m. Thalweg depth measurements (in the deepest part of channel), habitat classification, and midchannel substrate observations were made at tightly spaced intervals; whereas channel crosssections and shoreline-riparian stations for measuring or observing substrate, fish cover (concealment features), large woody debris, bank characteristics and riparian vegetation structure were spaced further apart. Thalweg (maximum) depth was measured at points evenly spaced every 0.4 ChW along these reaches to give profiles consisting of 100 measurements (150 in streams <2.5m wide). The tightly spaced depth measures allow calculation of indices of channel structural complexity, objective classification of channel units such as pools, and quantification of residual pool depth, pool volume, and total stream volume. Channel slope and sinuosity on non-wadeable rivers were estimated from 1:24,000-scale digital topographic maps.

In wadeable streams, wetted width was measured and substrate size and embeddedness were evaluated using a modified Wolman pebble count of 105 particles spaced systematically along 21 equally spaced cross-sections, in which individual particles were classified visually into seven sizeclasses plus bedrock, hardpan and other (e.g., organic material). The numbers of pieces of large woody debris in the bankfull channel were tallied in 12 size classes (3 length by 4 width classes) along the entire length of sample reaches. Channel incision and the dimensions of the wetted and bankfull stream channel were measured at 11 equally-spaced transects. Bank characteristics and areal cover of fish concealment features were visually assessed in 10 m long instream plots centered on transects, while riparian vegetation structure, presence of large (legacy) riparian trees, non-native (alien) riparian plants, and evidence of human disturbances (presence/absence and proximity) in 11 categories were visually assessed on adjacent 10 \times 10 m riparian plots on both banks. In addition, channel gradient (slope) in wadeable streams was measured to provide information necessary for calculating reach gradient residual pool volume and relative bed stability. In wadeable streams, crews used laser or hydrostatic levels for slopes <2.5%, and optionally were allowed to use hand-held clinometers in channels with slopes >2.5%. Compass bearing between stations were obtained for calculating channel sinuosity. Channel constraint and evidence of debris

torrents and major floods were assessed over the whole reach after the other components were completed. Discharge was measured by the velocity-area method at the time of sampling, or by other approximations if that method was not practicable (Peck et al. 2006; USEPA 2007). Twoperson crews typically completed NRSA habitat measurements in 1.5 to 4 hours of field time, though large, deep streams that were only marginally wadeable took up to several hours longer.

In non-wadeable rivers, NRSA field crews floating downstream in inflatable rafts, or in slower rivers small power boats, measured the longitudinal thalweg depth profile (approximated at mid-channel) using 7.5m telescoping survey rods or SONAR, at the same time tallying snags and off-channel habitats, classifying main channel habitat types, and characterizing mid-channel substrate by probing the bottom. At 11 littoral/riparian plots (each 10m wide x 20m long) spaced systematically and alternating sides along the river sample reach, field crews measured channel wetted width, bankfull channel dimensions, incision, channel constraint. They assessed near-shore, shoreline, and riparian physical habitat characteristics by measuring or observing littoral depths, riparian canopy cover, substrate, large woody debris, fish cover, bank characteristics, riparian vegetation structure, presence of large ("legacy") riparian trees, non-native riparian plants, and evidence of human activities. After all the thalweg and littoral/riparian measurements and observations were completed, the crews estimated the extent and type of channel constraint (see Peck et al. in press; USEPA 2007). Channel slope and sinuosity on non-wadeable rivers were estimated from 1:24,000 scale digital topographic maps.

See Kaufmann et al. (1999) for calculations of reach-scale summary metrics from field data, including mean channel dimensions, residual pool depth, bed particle size distribution, wood volume, riparian vegetation cover and complexity, and proximity-weighted indices of riparian human disturbances. See Faustini and Kaufmann (2007) for details on the calculation of geometric mean streambed particle diameter, and Kaufmann et al. (2008, 2009) for calculation of bed shear stress and relative bed stability which have been modified since published by Kaufmann et al. (1999), and Kaufmann and Faustini (2012) demonstrating the utility of EMAP and NRSA channel morphologic data to estimate transient storage and hydraulic retention in wadeable streams.

*2.*2 Quantifying the Precision of Physical Habitat Indicators

The absolute and relative precision of the physical habitat condition metrics used in NRSA are shown in Table 2, determined for most of the variables based on 2113 unique sites and repeat visits to a random subset of 197 of those sites. The RMS_{rep} expresses the precision or replicability of field measurements, quantifying the average variation in a measured value between same-season site revisits, pooled across all sites where measurements were repeated. We calculated RMS $_{\text{ren}}$ as the root-mean-square error of repeat visits during the same year, equivalent to the root mean-square error (RMSE) relative to the site means, as discussed Kaufmann et al., 1999 and Stoddard et al. (2005a). S/N is the ratio of variance among streams to that for repeat visits to the same stream as, described by Kaufmann et al. (1999).

The ability of a monitoring program to detect trends is sensitive to the spatial and temporal variation in the target indicators as well as the design choices for the network of sites and the timing and frequency of sampling. Sufficient temporal sampling of sites was not available to estimate all relevant components of variance for the entire U.S. However, Larsen et al (2004) examined the components of sampling variability for a number of the EMAP physical habitat variables including some of interest in this paper (residual depth, canopy cover, fine sediment, and large wood). Their analysis was based on evaluation on six Pacific Northwest surveys that included 392 stream reaches and 200 repeat visits. These surveys were conducted in Oregon and Washington from 1993 to 1999. Most were from one to three years in duration, but one survey lasted six years. They modeled the likelihood of detecting a 1–2% per year trend in the selected physical habitat
characteristics, if such a trend occurs, as a function of the duration of a survey. To calculate the number of years required to detect the defined trends in a monitoring network with a set number of sites, they set the detection probability at >80% with <5% probability of incorrectly asserting a trend if one is not present. We used the same survey data sets to duplicate their analysis for several variables not included in the Larsen et al. (2004) publication, including log transformed relative bed stability (*LRBS_BW5)* and riparian vegetation cover complexity (*XCMGW*, the combined cover of three layers of riparian woody vegetation); the results of that trend detection potential is summarized in Table 3.

3.0 Physical Habitat Condition Indicators in the NRSA

3.1 Relative Bed Stability and Excess Fines

Streambed characteristics (e.g., bedrock, cobbles, silt) are often cited as major controls on the species composition of macroinvertebrate, periphyton, and fish assemblages in streams (e.g., Hynes 1970, Cummins 1974, Platts et al. 1983, Barbour et al. 1999, Bryce et al., 2008, 2010). Along with bedform (e.g., riffles and pools), streambed particle size influences the hydraulic roughness and consequently the range of water velocities in a stream channel. It also influences the size range of interstices that provide living space and cover for macroinvertebrates and smaller vertebrates. Accumulations of fine substrate particles (excess fine sediments) fill the interstices of coarser bed materials, reducing habitat space and its availability for benthic fish and macroinvertebrates (Hawkins et al. 1983, Platts et al. 1983, Rinne 1988). In addition, these fine particles impede circulation of oxygenated water into hyporheic habitats reducing egg-to-emergence survival and growth of juvenile salmonids (Suttle et al. 2004). Streambed characteristics are often sensitive indicators of the effects of human activities on streams (MacDonald et al. 1991, Barbour et al. 1999, Kaufmann et al. 2009). Decreases in the mean particle size and increases in streambed fine sediments can destabilize stream channels (Wilcock 1997, 1998) and may indicate increases in the rates of upland erosion and sediment supply (Lisle 1982, Dietrich et al. 1989).

"Unscaled" measures of surficial streambed particle size, such as percent fines or D_{50} , can be useful descriptors of stream bed conditions. In a given stream, increases in percent fines or decreases in D₅₀ may result from anthropogenic increases in bank and hillslope erosion. However, a great deal of the variation in bed particle size among streams is natural: the result of differences in stream or river size, slope, and basin lithology. The power of streams to transport progressively larger sediment particles increases in direct proportion to the product of flow depth and slope. All else being equal, steep streams tend to have coarser beds than similar size streams on gentle slopes. Similarly, the larger of two streams flowing at the same slope will tend to have coarser bed material, because its deeper flow has more power to scour and transport fine particles downstream (Leopold et al. 1964, Morisawa 1968). For these reasons, we "scale" bed particle size metrics, expressing bed particle size in each stream as a deviation from that expected as a result of its size, power, and landscape setting (Kaufmann et al., 1999, 2008, 2009).

The scaled median streambed particle size is expressed as Relative Bed Stability (*RBS*), calculated as the ratio of the geometric mean diameter, D_{q} , divided by D_{cbf} , the critical diameter (maximum mobile diameter) at bankfull flow (Gordon et al., 1992), where D_g is based on systematic streambed particle sampling ("pebble counts") and D_{cbf} is based on the estimated streambed shear stress calculated from slope, channel dimensions, and hydraulic roughness during bankfull flow conditions

RBS is a measure of habitat stability for aquatic organisms as well as an indication of the potential for economic risk to streamside property and structures from stream channel movement. In many regions of the U.S.A, we may also be able to use *RBS* to infer whether sediment supply is augmented by upslope or bank erosion from anthropogenic or other disturbances, because it can indicate the degree of departure from a balance between sediment supply and transport. In

interpreting *RBS* on a regional scale, Kaufmann et al. (1999, 2009) argued that, over time, streams and rivers adjust sediment transport to match supply from natural weathering and delivery mechanisms driven by the natural disturbance regime, so that *RBS* in appropriately stratified regional reference sites should tend towards a range characteristic of the climate, lithology, and natural disturbance regime. Values of the *RBS* index either substantially lower (finer, more unstable streambeds) or higher (coarser, more stable streambeds) than those expected based on the range found in least-disturbed reference sites within an ecoregion are considered to be indicators of ecological stress.

Excess fine sediments can destabilize streambeds when the supply of sediments from the landscape exceeds the ability of the stream to move them downstream. This imbalance results from numerous human uses of the landscape, including agriculture, road building, construction and grazing. Lower-than-expected streambed stability may result either from high inputs of fine sediments (from erosion) or increases in flood magnitude or frequency (hydrologic alteration). When low *RBS* results from fine sediment inputs, stressful ecological conditions result from fine sediments filling in the habitat spaces between stream cobbles and boulders (Bryce et al. 2008, 2010). Instability (low *RBS*) resulting from hydrologic alteration can be a precursor to channel incision and arroyo formation (Kaufmann et al. 2009). Perhaps less well recognized, streams that have higher than expected streambed stability can also be considered stressed—very high bed stability is typified by hard, armored streambeds, such as those often found below dams where fine sediment flows are interrupted, or within channels where banks are highly altered. Values of *RBS* higher than reference expectations can indicate anthropogenic coarsening or armoring of streambeds, but streams containing substantial amounts of bedrock may also have very high *RBS*, and at this time it is difficult to determine the role of human alteration in stream coarsening on a national scale. For this reason, NRSA reported only on the "low end" of *RBS* relative to reference conditions, generally indicating stream bed excess fine sediments or augmented stormflows associated with human disturbance of stream drainages and riparian zones.

Precision of Sediment and Bed Stability Measurements – The geometric mean bed particle diameter (*D*gm) and *RBS* varied over 8 orders of magnitude in the NRSA surveys. Because of this wide variation and the fact that both exhibit repeat-visit variation that is proportional to their magnitude at individual streams, it is useful and necessary to log transform these variables (*LSUB_DMM* and LRBS_g08). The RMS_{rep} of LSUB_DMM in wadeable streams of the EMAP-W survey was 0.246, similar to that reported by Faustini and Kaufmann (2007) for EMAP-W (0.21). For a $D_{\text{cm}} = \gamma v^2$ mm, the log-based RMSrep of 0.246 translates to an asymmetrical 1SD error bound of 0.57*y* to 1.76*y* mm. The RMS_{rep} of *LRBS_g08* in NRSA wadeable streams was 0.48, approximately 6% of its observed range, but less precise (surprisingly) than that for EMAP-W (RMS $_{\text{rep}}$ = 0.365). The logbased RMS_{rep} of 0.48 for NRSA LRBS_g08 translates to an asymmetrical error bound of 0.33*y* to 3.0*y* around an untransformed *RBS* value of *"y"* (Table 2). Compared with the high S/N ratio for *LSUB_DMM* in NRSA (12.4 for wadeable+boatable waters), relative precision for *LRBS_g08* was lower (S/N=5.0), reflecting the reduction in total variance when a large component of natural variability is "modeled out" by scaling for channel gradient, water depth, and channel roughness. Nevertheless, the relative precision of LRBS_g08 is moderately high and easily adequate to make it a useful variable in regional and national assessments (Kaufmann et al. 1999, 2008, Faustini and Kaufmann 2007). The transformation of the unscaled geometric mean bed particle diameter D_{cm} to the ratio *RBS* by dividing by the critical diameter reduced the within-region variation by accounting for some natural controlling factors. As a result, we feel that the scaled variable helps to reveal alteration of bed particle size and mobility from anthropogenic erosion and sedimentation (Kaufmann et al. 2008, 2009).

We have examined the components of variability of *LRBS* based on earlier surveys and modeled its potential utility in trend detection in the Pacific Northwest region of the U.S. with the same data and procedures as used by Larsen et al. (2004), in which all methods were the same as used in EMAP-W and WSA except that bed substrate mean diameter data used by Larsen et al. was determined based on 55, rather than 105 particles. (NRSA data differed from data used in that analysis by using laser levels rather than hand-held clinometers to measure wadeable stream slopes <2/5%) That analysis showed that a 50-site monitoring program could detect a subtle trend in *LRBS_BW5* of 2% per year within 8 years, if sites were visited every year (Table 3).

3.2 Instream Habitat Cover Complexity

Although the precise mechanisms are not completely understood, the most diverse fish and macroinvertebrate assemblages are usually found in streams that have complex mixtures of habitat features: large wood, boulders, undercut banks, tree roots, etc. (see Kovalenko et al. 2011). When other needs are met, complex habitat with abundant cover should generally support greater biodiversity than simple habitats that lack cover (Gorman and Karr 1978, Benson and Magnuson 1992). Human use of streams and riparian areas often results in the simplification of this habitat, with potential effects on biotic integrity (Kovalenko et al., 2011). For this assessment, we use a measure (XFC_NAT in Kaufmann et al., 1999) that sums the amount of instream habitat consisting of undercut banks, boulders, large pieces of wood, brush, and cover from overhanging vegetation within a meter of the water surface, all of which were estimated visually by NRSA field crews.

Quantifying Habitat Complexity – Habitat complexity is difficult to quantify, and could be quantified or approximated by a wide variety of measures. The NRSA Physical Habitat protocols provide estimates for nearly all of the following components of complexity identified during EPA's 1992 stream monitoring workshop (Kaufmann 1993):

• Habitat type and distribution (e.g., Bisson et al. 1982, O'Neill and Abrahams 1984, Frissell et al. 1986, Hankin and Reeves 1988, Hawkins et al. 1993, Montgomery and Buffington 1993, 1997, 1998).

• Large wood count and size (e.g., Harmon et al. 1986, Robison and Beschta 1989, Peck et al. 2006).

• In-channel cover: Percentage areal cover of fish concealment features, including undercut banks, overhanging vegetation, large wood, boulders (Hankin and Reeves 1988, Kaufmann and Whittier 1997, Peck et al. 2006)

• Residual pools, channel complexity, hydraulic roughness(e.g., Kaufmann 1987a, b, Lisle 1987, Stack and Beschta, 1989; Lisle and Hilton 1992, Robison and Kaufmann 1994, Kaufmann et al. 1999, Kaufmann et al. 2008, Kiem et al., 2002; Kaufmann et al. 2011)

• Width and depth variance, bank sinuosity (Kaufmann 1987a, Moore and Gregory 1988, Kaufmann et al. 1999, Madej 1999, 2001, Kaufmann et al. 2008, Mossop and Bradford 2006; Pearsons and Temple, 2007, 2010; Kaufmann and Faustini, 2011).

Residual depth is a measure of habitat volume, but also serves as one of the indicators of channel habitat complexity, particularly when expressed as a deviation from reference expectations, including the influences of basin size. A stream with more complex bottom profile will have greater residual depth than one of similar drainage area, discharge, and slope that lacks that complexity (Kaufmann 1987a). Conversely, between two streams of equal discharge and slope, the one with greater residual depth (i.e., larger, more abundant residual pools) will have greater variation in cross-sectional area, slope, and substrate size. A related measure of the complexity of channel morphology is the coefficient of variation in thalweg depth, calculated entirely from the thalweg depth profile (*SDDEPTH / XDEPTH*). The thalweg profile is a systematic survey of depth in the stream channel along the path of maximum depth ("thalweg"). In addition to measures of channel morphometric complexity, EMAP habitat protocols measure in-channel large wood (sometimes called "large woody debris" or simply "LWD"), and several estimates of the areal cover of various types of fish and macroinvertebrate "cover" or concealment features. The large wood metrics include counts of wood pieces per 100 m of bankfull channel and estimates of large wood volume in the sample reach expressed in cubic meters of wood per square meter of bankfull channel. The "fish cover" variables are visual estimates of the areal cover of single or combined types of habitat features.

NRSA required a general summary metric as a holistic indicator of many aspects of habitat complexity, so used the metric *XFC_NAT*, summing the areal cover from large wood, brush, overhanging vegetation, live trees and roots, boulders, rock ledges, and undercut banks in the wetted stream channel. Habitat complexity and the abundance of particular types of habitat features differ naturally with stream size, slope, lithology, flow regime, and potential natural vegetation. For example, boulder cover will not occur naturally in streams draining deep deposits of loess or alluvium that do not contain large rocks. Similarly, large wood will not be found naturally in streams located in regions where riparian or upland trees do not grow naturally. Though the combined cover index *XFC_NAT* partially overcomes these differences, we set stream-specific expectations for habitat complexity metrics in NRSA based on region-specific reference sites and further refined them as a function of geoclimatic controls.

Precision of Habitat Complexity Measures – The instream habitat complexity index *XFC_NAT* ranged from 0 to 2.3, or 0% to 230% in NRSA, expressing the combined areal cover of the five cover elements contributing to its sum. The RMSrep of Log(0.01+*XFC_NAT)* was 0.24, meaning that an XFC_NAT value of 10% cover at a single stream site has a $+1.0$ RMS_{rep} error bound of 6% to 17% (Table 2) S/N was relatively low for this indicator (1.87), though higher in wadeable streams (2.29) than in boatable rivers (1.22). Despite its relatively low S/N , the RMS_{rep} for LXFC_LWD was 10% of the observed range of XFC_NAT. It was retained as a habitat complexity indicator because it contains biologically relevant information not available in other metrics, showed moderate responsiveness to human disturbances, and has precision adequate to discern relatively large differences in habitat complexity.

3.3 Riparian Vegetation

Quantifying Riparian Vegetation Cover Complexity – The importance of riparian vegetation to channel structure, cover, shading, inputs of nutrients and large wood, and as a wildlife corridor and buffer against anthropogenic disturbance is well recognized (Naiman et al. 1988, Gregory et al. 1991). Riparian vegetation not only moderates stream temperatures through shading, but also increases bank stability and the potential for inputs of coarse and fine particulate organic material. Organic inputs from riparian vegetation become food for stream organisms and provide structure that creates and maintains complex channel habitat.

The presence of a complex, multi-layered vegetation corridor along streams and rivers is an indicator of how well the stream network is buffered against sources of stress in the watershed. Intact riparian areas can help reduce nutrient and sediment runoff from the surrounding landscape, prevent streambank erosion, provide shade to reduce water temperature, and provide leaf litter and large wood that serve as food and habitat for stream organisms (Gregory et al., 1991). The presence of large, mature canopy trees in the riparian corridor reflects its longevity, whereas the presence of smaller woody vegetation typically indicates that riparian vegetation is reproducing, and suggests the potential for future sustainability of the riparian corridor.

NRSA evaluated the cover and complexity of riparian vegetation based on the metric *XCMGW*, which is calculated from visual estimates made by field crews of the areal cover and type of vegetation in three layers: the ground layer (<0.5m), med-layer (0.5-5.0 m) and upper layer (>5.0 m). The separate measures of large and small diameter trees, woody and non-woody mid-layer vegetation, and woody and non-woody ground cover are all visual estimates of areal cover. *XCMGW* sums the cover of *woody* vegetation summed over these three vegetation layers, expressing both the abundance of vegetation cover and its structural complexity. Its theoretical maximum is 3.0 if there is 100% cover in each of the three vegetation layers. *XCMGW* gives an indication of the longevity and sustainability of perennial vegetation in the riparian corridor (Kaufmann et al. 1999).

Precision of Riparian Vegetation index – XCMGW ranged from 0 to 2.8 (280% cover), with RMS_{rep} of $Log(0.01+XCMGW) = 0.146$ (Table 2), meaning that an XCMGW value of 10% at a single stream site has a $+1.0$ RMS_{rep} error bound of 7% to 14%. Its S/N ratio was 9.38, indicating very good potential for discerning differences among sites. We examined the components of variability of *XCMGW* and modeled its potential utility in trend detection in the Pacific Northwest region of the U.S. with the same data and procedures as used by Larsen et al. (2004). Based on that analysis, a 50-site monitoring program could detect a subtle trend in *XCMGW* of 2% per year within 8 years, if sites were visited every year (Table 3).

3.4 Riparian Human Disturbances

Agriculture, roads, buildings, and other evidence of human activities in and near the stream and river channel may exert stress on aquatic ecosystems and may also serve as indicators of overall anthropogenic stress. EPA's 1992 stream monitoring workshop recommended field assessment of the frequency and extent of both in-channel and near-channel human activities and disturbances (Kaufmann 1993). The vulnerability of the stream network to potentially detrimental human activities increases with the proximity of those activities to the streams themselves. NRSA follow Stoddard et al. (2005b) and U.S. EPA (2006) in using a direct measure of riparian human disturbance that tallies 11 specific forms of human activities and disturbances (walls, dikes, revetments or dams; buildings; pavement or cleared lots; roads or railroads; influent or effluent pipes; landfills or trash; parks or lawns; row crop agriculture; pasture or rangeland; logging; and mining) at 22 separate locations along the stream reach, and weights them according to how close to the channel they are observed (*W1_HALL* in Kaufmann et al. 1999). Observations within the stream or on its banks are weighted by 1.5, those within the 10 \times 10 meter plots are weighted by 1.0, and those visible beyond the plots are weighted by 0.5. The index *W1* HALL ranged from 0 (no observed disturbance) to \sim 7 (e.g., equivalent to four or 5 types of disturbance observed in the stream, throughout the reach; or seven types observed within all 22 riparian plots bounding the stream reach). Although direct human activities certainly affect riparian vegetation complexity and layering measured by the Riparian Vegetation Index (previous paragraph), the Riparian Disturbance Index is more encompassing, and differs by being a *direct* measure of observable human activities that are presently or potentially detrimental to streams.

Precision of Riparian Disturbance Indicators – W1_HALL ranged from 0 to 7.3 in NARS. The precision of the weighted human disturbance indicator *W1_HALL* was proportional to the level of disturbance. The RMS_{rep} of log(0.1+W1_HALL) was 0.186 (Table 2), meaning that a W1_HALL value of 1.0 at a single stream site has a $+1.0$ RMS_{rep} error bound of 0.65 to 1.53. The relative precision of Log(0.1+*W1_HALL*) was moderate (S/N=5.18)

4.0 Estimating Reference Condition for Physical Habitat

4.1 Reference Site Screening and Anthropogenic Disturbance Classifications

As part of the routine application of its field and GIS protocols, NRSA obtained various measures of human disturbance associated with each site and its catchment. Site scale indicators of human disturbance included evidence of various human activities including nearby roads, riprap, agricultural activities, riparian vegetation disturbance, etc., as detailed by Kaufmann et al. (1999). These indicators of local scale disturbance were used in combination with water chemistry (Chloride, Total Phosphorus, Total Nitrogen, and Turbidity), as described by Herlihy et al. (2008), to screen

probability and hand-picked sites and designate them as least- moderately-, and highly-disturbed, relative to other sites within each of the regions of NRSA. In addition, we used basin and sub-basin row crop and urban land use percentages and the density of dams and impoundments as described in the reference technical section to rank sites by disturbance categories, as shown in Table 4. To avoid circularity, we did not use any field measures of sediment, in-channel habitat complexity, or riparian vegetation to screen least-disturbed sites used to estimate reference condition for excess streambed fining, instream fish cover, and riparian vegetation. Nor did we use such measures in defining levels of disturbance to use in examining the associations of these habitat metrics with human disturbances. We did, however, use field observations of the level and proximity of streamside human activities in screening reference sites and defining levels of disturbance for evaluating indicator responsiveness. In this article, the designation "R" refers to least-disturbed ("reference") sites; "M" to moderately-disturbed sites, and "D" to the most-disturbed sites within each of the nine aggregate ecoregions discussed herein. We defined these site disturbance categories independent of the habitat indicators we evaluate in this article (other than riparian human disturbances), allowing an assessment of fluvial habitat response to a gradient of human activities and disturbances.

4.2 Modeling Expected Reference Values of the Indicators

For LRBS, we modeled expected values based on the distribution of LRBS in reference sites within regions or groups of regions. In some regions boatable and wadeable rivers and streams are modeled separately; in others they are combined. Where possible, we used regression models in which W1 Hall (human disturbance) is set to zero in regressions of LRBS=f(W1 Hall) within reference sites only (RMD_PHAB=R). In these cases, the adjusted mean of the reference distribution is defined as the y-intercept of these regressions and the SD about the adjusted reference mean is defined as the RMSE of those regressions. Condition classes were defined based on normal approximation of the 5th and 25th percentiles of the actual or adjusted reference distributions. The definition of "Poor" condition was set as those sites with LRBS < the reference mean LRBS minus $1.65(SD_{ref})$. Sites in "Good" condition with respect to this indicator were those with LRBS> the reference mean LRBS minus $0.67(SD_{ref})$.

For instream fish cover complexity, we estimated expected XFC_NAT based on multiple linear regression models predicting Log₁₀(XFC_NAT) in reference sites from geoclimatic controlling factors within regions or aggregated regions. Because there is a gradient of human disturbance within the set of reference sites in all the regions considered, and it was correlated with XFC_NAT, we also incorporated field measures of human disturbance into the regressions. Site-specific expected ("E") values of XFC_NAT were then calculated by setting human the disturbance metric values to very low values (but never lower than observed among the reference). We then calculated observed/expected (O/E) values of XFC_NAT and examined their distribution among reference sites. Because we had modeled-out disturbance to some extent in our calculation of E values, the distribution of O/E in reference sites did not necessarily have a mean of 1/1 (Log=0), although means were very close to 1/1. We set expectations of the O/E values based on the mean and SD of the regional reference distributions, analogous to that described for LRBS in the previous paragraph.

For riparian condition (XCMGW transformed as L_{x} cmgw= Log₁₀(0.01+XCMGW) we estimated expected condition based on simple regional reference site distributions or regression models in which W1_Hall was set to zero in regressions prediction L_xcmgw as a function of W1_Hall within the subset of reference sites (RMD_PHAB=R). --- The adjusted mean L_xcmgw for reference sites was defined as the y-intercept and the SD about the reference mean is defined as the RMSE of those regressions.

We did not base thresholds of riparian disturbance on the reference distributions, as was done for sediment, habitat complexity and riparian vegetation condition. Rather, the classes for riparian disturbance were set using the same judgement-based criteria for all regions. W1_HALL, the database variable name for this indicator, is a direct measure of human disturbance "pressure" – unlike the other habitat indicators, which are actually measures of habitat response to human disturbance pressures. It is very difficult to define reference sites without screening sites based on these human disturbance tallies (i.e., W1_HALL). For this reason, we took a different approach for setting riparian disturbance thresholds, defining low disturbance sites as those with W1_HALL <0.33 and high riparian disturbance sites as those with W1_HALL >1.5; we applied these same thresholds in all ecoregions. A value of 1.5 for a stream means, for example, that at 22 locations along the stream the field crews found an average of one of 11 types of human disturbance within the stream or its immediate banks. A value of 0.33 means that, on average, one type of human disturbance was observed at one-third of the 22 riparian plots along a sample stream or river.

5.0 Response of the Physical Habitat Indicators to Human Disturbance

The Sedimentation and Riparian Vegetation indicators, LRBS and XCMGW showed modest to strong negative response to human disturbance in most regions and aggregations of regions, as illustrated by t-values (+2.11-12.24) comparing differences in means of Reference minus Disturbed sites (Table 5). However, the strength of sediment and riparian vegetation associations with human disturbance tended to be slightly stronger for sediments and much stronger for riparian vegetation in wadeable *versus* boatable sites (Table 5, and Figures 1 and 2).

Except for the weak contrary response in the Eastern Highlands ($t = -1.26$), the instream habitat complexity indicator showed moderate to moderate response to human disturbance, with t-values ranging from +2.13 to +4.25 (Table 5). As for the other habitat indicators, associations were in most cases stronger for wadeable, *versus* boatable sites.

Because the field-obtained measures of riparian disturbance used in the NRSA are themselves direct indicators of human disturbance, and were used to screen reference sites, we illustrate the relationship of *W1_HALL* to the human disturbance gradient in Figures 1 and 2 only to compare their relative magnitudes among disturbed and undisturbed streams in the various regions of the U.S.

6.0 Literature Cited

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. EPA/841- B-99-002, U.S. Environmental Protection Agency, Washington, D.C.

Benson, B. J. and J. J. Magnuson. 1992. Spatial heterogeneity of littoral fish assemblages in lakes: relation to species diversity and habitat structure. Canadian Journal of Fisheries and Aquatic Sciences 49:1493-1500.

Bisson, P. A., J. L. Nielsen, R. A. Palmason, and L. E. Grove. 1982. A system of naming habitat types in small streams, with examples of habitat utilization by salmonids during low stream flow. Pages 62-73 In N. B. Armantrout, [editor]. Acquisition and utilization of aquatic habitat inventory information. Symposium Proceedings, October 28-30, 1981, Portland, Oregon. The Hague Publishing, Billings, Montana.

Bryce, S.A., G.A. Lomnicky, P.R. Kaufmann, L.S. McAllister, and T.L. Ernst. 2008. Development of Biologically-Based Sediment Criteria in Mountain Streams of the Western United States. N. Am. J. Fish. Manage. 28: 28:1714-1724.

Bryce, S.A., Lomnicky, G.A, and P.R. Kaufmann. 2010. Protecting Sediment-Sensitive Aquatic Species in Mountain Streams through the Application of Biologically-Based Criteria Streambed Sediment Criteria. J. North American Benthological Soc. 29(2):657-672.

Buffington, J. M. and D. R. Montgomery. 1999a. Effects of hydraulic roughness on surface textures of gravel-bed rivers. Water Resources Research 35:3507-3521.

Buffington, J. M. and D. R. Montgomery. 1999b. Effects of sediment supply on surface textures of gravel-bed rivers. Water Resources Research 35:3523-3530.

Cummins, K. W. 1974. Structure and function of stream ecosystems. BioScience 24:631-641. Dietrich, W. E., J. W. Kirchner, H. Ikeda, and F. Iseya. 1989. Sediment supply and the development of the coarse surface layer in gravel bed rivers. Nature 340:215-217.

Dingman, S. L. 1984. Fluvial Hydrology. W.H. Freeman, New York.

Dunne, T. and L. B. Leopold. 1978. Water in environmental planning. W. H. Freeman and Co., New York

Faustini, J. M. and P. R. Kaufmann. 2007. Adequacy of Visually Classified Particle Count Statistics From Regional Stream Habitat Surveys1. Journal of the American Water Resources Association 43:1293-1315.

Fitzpatrick, F. A., I. R. Waite, P. J. D'Arconte, M. R. Meador, M. A. Maupin, and M. E. Gurtz. 1998. Revised methods for characterizing stream habitat in the National Water-Quality Assessment Program. Water-Resources Investigations Report 98-4052, U.S. Geological Survey Reston, Virginia. Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environmental Management 10:199- 214.

Gordon, N. D., T. A. McMahon, and B. L. Finlayson. 1992. Stream hydrology, an introduction for ecologists. John Wiley & Sons, New York.

Gorman, O. T. and J. R. Karr. 1978. Habitat structure and stream fish communities. Ecology 59:507- 515.

Gregory, S. V., F. J. Swanson, W. A. McKee, and K. W. Cummins. 1991. An Ecosystem Perspective of Riparian Zones. BioScience 41:540-551.

Hankin, D. G. and G. H. Reeves. 1988. Estimating total fish abundance and total habitat area in small streams based on visual estimation methods. Canadian Journal of Fisheries and Aquatic Sciences 45:834-844.

Harmon, M. E., J. F. Franklin, F. J. Swanson, P. Sollins, S. V. Gregory, J. D. Lattin, N. H. Anderson, S. P. Cline, N. G. Aumen, J. R. Sedell, G. W. Lienkamper, K. Cormack Jr., and K. W. Cummins. 1986. Ecology of Coarse Woody Debris in Temperate Ecosystems. Advances in Ecological Research 15:133-302.

Harrelson, C. C., C. L. Rawlins, and J. P. Potyondy. 1994. Stream channel reference sites: an illustrated guide to field technique. General Tech. Rep. RM-245, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

Hawkins, C. P., J. L. Kershner, P. A. Bisson, M. D. Bryant, L. M. Decker, S. V. Gregory, D. A. McCullough, C. K. Overton, G. H. Reeves, R. J. Steedman, and M. K. Young. 1993. A Hierarchical Approach to Classifying Stream Habitat Features. Fisheries 18:3-12.

Hawkins, C. P., M. L. Murphy, and N. H. Anderson. 1983. Density of fish and salamanders in relation to riparian canopy and physical habitat in streams of the northwestern United States. Canadian Journal of Fisheries and Aquatic Sciences 40:1173-1186.

Herlihy, A.T., S.G. Paulsen., J. Van Sickle, J.L. Stoddard. C.P. Hawkins, L.L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. J. N. Am. Benthol. Soc. 27(4):860–877.

Hynes, H. B. N. 1970. The ecology of running waters. Unversity of Toronto Press, Toronto, Ontario, Canada.

Kappesser, G. B. 2002. A riffle stability index to evaluate sediment loading to streams. Journal of the American Water Resources Association 38:1069-1081.

Kaufmann, P. R. 1987a. Channel morphology and hydraulic characteristics of torrent-impacted forest streams in the Oregon Coast Range, U.S.A. Ph.D. Dissertation. Oregon State University, Corvallis, Oregon.

Kaufmann, P. R. 1987b. Slackwater habitat in torrent-impacted streams. Pages 407-408 In R. L. Beschta, T. Blinn, G. E. Grant, F. J. Swanson, and G. E. Ice, [editors]. Erosion and Sedimentation in the Pacific Rim. International Association of Hydrologic Science, Pub. No. 165, Proceedings of an International Symposium, August 3-7, 1986, Ore. State Univ., Corvallis, OR. International Association of Hydrologic Science.

Kaufmann, P. R. 1993. Physical habitat. Pages 59-69 In R. M. Hughes. Stream Indicator and Design Workshop. EPA 600/R-93/138, U.S. Environmental Protection Agency, Office of Research and Development, Corvallis, Oregon.

Kaufmann, P. R., J. M. Faustini, D. P. Larsen, and M. A. Shirazi. 2008. A roughness-corrected index of relative bed stability for regional stream surveys. Geomorphology 199:150-170.

Kaufmann, P.R., D.P. Larsen, and J.M. Faustini, 2009. Bed Stability and Sedimentation Associated With Human Disturbances in Pacific Northwest Streams. J. Am. Water Resources Assoc. 45(2):434- 459.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA/620/R-99/003, U.S. Environmental Protection Agency, Washington, D.C.

Kaufmann, P. R. and T. R. Whittier. 1997. Habitat Assessment. Pages 5-1 to 5-26 In J. R. Baker, D. V. Peck, and D. W. Sutton. Environmental Monitoring and Assessment Program -Surface Waters: Field Operations Manual for Lakes. EPA/620/R-97/001, U.S. Environmental Protection Agency, Washington, D.C.

Kaufmann, P. R. & J. M. Faustini, 2011. Simple measures of channel habitat complexity predict transient hydraulic storage in streams. Hydrobiologia. doi:10.1007/s10750-011-0841-y.

Keim, R. F., A. E. Skaugset & D. S. Bateman, 2002. Physical aquatic habitat II. Pools and cover affected by large woody debris in three western Oregon streams. North American Journal of Fisheries Management 22: 151–164.

Kovalenko, K.E., S.M. Thomaz, and D.M. Warfe. 2012. Habitat complexity: approaches and future directions – editorial review. Hydrobiologia 685:1–17. DOI 10.1007/s10750-011-0974-z

Larsen, D. P., P. R. Kaufmann, T. M. Kincaid, and N. S. Urquhart. 2004. Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. Canadian Journal of Fisheries and Aquatic Sciences 61:283-291.

Lazorchak, J. M., D. J. Klemm, and D. V. e. Peck. 1998. Environmental Monitoring and Assessment Program-Surface Waters: field operations and methods for measuring the ecological condition of wadeable streams. EPA/620/R-94/004F, U.S. Environmental Protection Agency, Washington, D.C.

Leopold, L. B. 1994. A View of the River. Harvard University Press, Cambridge, Massachusetts. Leopold, L. B., M. G. Wolman, and J. P. Miller. 1964. Fluvial processes in geomorphology. W.H. Freeman, San Francisco.

Lisle, T. E. 1982. Effects of aggradation and degradation on riffle-pool morphology in natural gravel channels, northwestern California. Water Resources Research 18:643-1651.

Lisle, T. E. 1987. Using "residual depths" to monitor pool depths independently of discharge. Research Note PSW-394, USDA Forest Service, Pacific Southwest Forest and Range Experimental Station, Berkeley, California.

Lisle, T. E. and S. Hilton. 1992. The volume of fine sediment in pools: an index of sediment supply in gravel-bed streams. Water Resources Bulletin 28:371-383.

MacDonald, L. H., A. W. Smart, and R. C. Wismar. 1991. Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska. EPA 910/9-91-001, U.S. Environmental Protection Agency, Region X, Seattle, Washington.

Madej, M. A. 2001. Development of channel organization and roughness following sediment pulses in single-thread, gravel bed rivers. Water Resources Research 37:2259-2272.

Montgomery, D. R. and J. M. Buffington. 1993. Channel classification, prediction of channel response, and assessment of channel condition. Washington State Timber/Fish/Wildlife Agreement, Report TFW-SH10-93-002, Department of Natural Resources, Olympia, Washington.

Montgomery, D. R. and J. M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. Geological Society of America Bulletin 109.

Montgomery, D. R. and J. M. Buffington. 1998. Channel processes, classification, and response. Pages 13-42 In R. Naiman and R. Bilby, [editors]. River Ecology and Management. Springer-Verlag, New York.

Moore, K. M. S. and S. V. Gregory. 1988. Summer habitat utilization and ecology of cutthroat trout fry (*Salmo clarki*) in Cascade mountain streams. Canadian Journal of Fisheries and Aquatic Sciences 45:1921-1930.

Morisawa, M. 1968. Streams, their dynamics and morphology. McGraw-Hill Book Company, New York.

Mossop, B. & M. J. Bradford, 2006. Using thalweg profiling to assess and monitor juvenile salmon (*Oncorhynchus spp*.) habitat in small streams. Canadian Journal of Fisheries and Aquatic Sciences 63: 1515–1525.

Naiman, R. J., H. Decamps, J. Pastor, and C. A. Johnston. 1988. The potential importance of boundaries to fluvial ecosystems. Journal of the North American Benthological Society 7:289-306.

O'Neill, M. P. and A. D. Abrahams. 1984. Objective identification of pools and riffles. Water Resources Research 20:921-926.

Paulsen, S.G., A. Mayio, D.V. Peck, J.L. Stoddard, E.Tarquinio, S.M. Holdsworth, J. Van Sickle, L.L. Yuan, C.P. Hawkins, A.T. Herlihy, P.R. Kaufmann, M.T. Barbour, D.P. Larsen, and A.R. Olsen. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. J. N. Am. Benthological Soc. 27(4):812–821.

Pearsons, T. N. & G. M. Temple, 2007. Impacts of early stages of salmon supplementation and reintroduction programs on three trout species. North American Journal of Fisheries Management 27: 1–20.

Pearsons, T. N. & G. M. Temple, 2010. Changes to Rainbow Trout abundance and salmonid biomass in a Washington watershed as related to hatchery salmon supplementation. Transactions of the American Fisheries Society 139: 502–520.

Peck, D. V., D. K. Averill, A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, M. R. Cappaert, T. Magee, and P. A. Monaco. in press. Environmental Monitoring and Assessment Program: Surface Waters Western Pilot Study—field operations manual for nonwadeable streams. EPA 620/ R-xx/xxx, U.S. Environmental Protection Agency, Washington, D.C.

Peck, D. V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. Magee, and M. R. Cappaert. 2006. Environmental Monitoring and Assessment Program: Surface Waters Western Pilot Study—field operations manual for wadeable streams. EPA/620/R-06/003, U.S. Environmental Protection Agency, Washington, D.C.

Platts, W. S., W. F. Megahan, and G. W. Minshall. 1983. Methods for evaluating stream, riparian, and biotic conditions. General Technical Report INT-138, U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah.

Rinne, J. 1988. Effects of livestock grazing exclosure on aquatic macroinvertebrates in a montane stream, New Mexico. Great Basin Naturalist 48:146-153.

Robison, E. G. and R. L. Beschta. 1989. Estimating stream cross sectional area from wetted width and thalweg depth. Physical Geography 10:190-198.

Robison, E. G. and P. R. Kaufmann. 1994. Evaluating two objective techniques to define pools in small streams. Pages 659-668 In R. A. Marston and V. A. Hasfurther, [editors]. Effects of Human Induced Changes on Hydrologic Systems. Summer Symposium proceedings, American Water Resources Association,. June 26-29, 1994, Jackson Hole, Wyo.

Stack, W. R. and R. L. Beschta. 1989. Factors influencing pool morphology in Oregon coastal streams. Pages 401-411 In W. W. Woessner and D. F. Potts, [editors]. Headwaters Hydrology Symposium. American Water Resources Association.

Stoddard, J. L., D. V. Peck, A. R. Olsen, D. P. Larsen, J. Van Sickle, C. P. Hawkins, R. M. Hughes, T. R. Whittier, G. Lomnicky, A. T. Herlihy, P. R. Kaufmann, S. A. Peterson, P. L. Ringold, S. G. Paulsen, and R. Blair. 2005a. Environmental Monitoring and Assessment Program (EMAP): western streams and rivers statistical summary. EPA 620/R-05/006, U.S. Environmental Protection Agency, Washington, D.C.

Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. P. Larsen, G. Lomnicky, A. R. Olsen, S. A. Peterson, P. L. Ringold, and T. R. Whittier. 2005b. An ecological assessment of western streams and rivers. EPA 620/R-05/005, U.S. Environmental Protection Agency, Washington, D.C.

Suttle, K. B., M. E. Power, J. M. Levine, and C. McNeely. 2004. How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. Ecological Applications 14:969–974. U.S. EPA. 2004. Wadeable Streams Assessment: field operations manual. EPA/841/B-04/004, U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA. 2006. Wadeable Streams Assessment: a collaborative survey of the Nation's streams. EPA/641/B-06/002, U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA 2007. National Rivers and Streams Assessment: Field Operations Manual EPA-841-B-07- 009. U.S. Environmental Protection Agency, Washington, D.C.

Wilcock, P. R. 1997. The components of fractional transport rate. Water Resources Research 33:247-258.

Wilcock, P. R. 1998. Two-fraction model of initial sediment motion in gravel-bed rivers. Science 280:410-412.

Background Information

Fish assemblages in streams and rivers offer several unique advantages to assess ecological condition, based on their mobility, longevity, trophic relationships, and socioeconomic importance [\(Barbour et al. 1999,](#page-135-0) [Roset et al. 2007\)](#page-136-0). For fish assemblages, assessing ecological condition has generally been based the development and use of multi-metric indices (MMIs), which are derivations of the original Index of Biotic Integrity (IBI) developed by Karr and others [\(Karr 1981,](#page-135-1) [Karr et al.](#page-135-2) [1986,](#page-135-2) [Karr 1991,](#page-135-3) [1999,](#page-135-4) [Karr and Chu 2000\)](#page-135-5). There are numerous examples of MMIs developed for fish assemblages in smaller streams [\(e.g., Bramblett et al. 2005\)](#page-135-6) as well as for larger rivers [\(Lyons](#page-136-1) [et al. 2001,](#page-136-1) [Emery et al. 2003,](#page-135-7) [Mebane et al. 2003,](#page-136-2) [Pearson et al. 2011\)](#page-136-3).

Recently, MMIs for fish assemblages have been developed based on applying the techniques used to develop predictive models of taxonomic composition (e.g., [Hawkins 2006,](#page-135-8) [Meador and Carlisle](#page-136-4) [2009\)](#page-136-4) to metrics instead of taxa [\(Pont et al. 2007,](#page-136-5) [Pont et al. 2009\)](#page-136-6). This approach essentially provides an estimate of expected condition (in terms of metric values) at individual sites, rather than utilizing a set of regional reference sites to define expected values for a particular metric, Hawkins et al. [\(2010a\)](#page-135-9) concluded that the combined approach resulted in MMIs that performed better in terms of their ability to discern deviation from expected condition

For NRSA, we developed multimetric indices for fish assemblages (FMMIs) using the combined approach (modeling expected values of metrics). We developed separate FMMIs for each of the three climatic regions (Eastern Highlands, Plains and Lowlands, and the West; Figure D-1).

Figure D-1. Aggregated Omernik ecoregions used for NRSA fish MMI development. Separate MMIs were developed for each of the three climatic regions.

METHODS

2.1 Field methods

Collection methods for fish are described in the NRSA field operations manual [\(USEPA 2009\)](#page-137-0). Three variants of the basic sampling protocol (using electrofishing) were used depending on the width of the stream and whether or not it was wadeable. For wadeable streams less than 12.5 m wide, 40 channel widths were sampled for fish. For larger wadeable streams (> 12.5 m wide), 500 m or 20 channel widths were sampled (whichever was longer). For non-wadeable streams and rivers, at least 20 channel widths were sampled. At large wadeable and non-wadeable sites, sampling continued past the established reach length until 500 individuals were collected.

2.2 Counting, Taxonomy, and Autecology

Fish were tallied and identified in the field, then released alive unless used for fish tissue or vouchers Voucher specimens were collected if field identification could not be accomplished. Voucher samples were also collected at 10% of sites for each taxonomist. All names submitted on field data forms were reviewed and revised when necessary to create a listing of nationally consistent common and scientific names. Where possible, taxonomic names (common and scientific) were based on Nelson et al. [\(2004\)](#page-136-7). The online database FishBase (http://www.fishbase.org) served as a secondary source of taxonomic names. In rare cases, a journal article of a newly described species was used. Collection maps for each taxon were prepared and compared to published maps in Page and Burr [\(1991\)](#page-136-8) A total of 631 unique taxa were identified, excluding unknowns, hybrids, and amphibians.

Each taxon was characterized for several different autecological traits, based on available sources of published information. Traits included habitat guilds (lotic habitat and temperature), trophic guild, reproductive guild, migration pattern, and tolerance to human disturbance. A list of all fish taxa and their associated autecological assignments are available as a tab-delimited data file that will be available on the NRSA website in December 2012

Assignment of native status were made at the scale of 8-digit Hydrologic Unit Codes (HUC). Published sources (USGS Nonindigenous Species database and NatureServe) were used as the basis for assigning a taxon collected at a particular site as being native or introduced.

Because fish collected at a site cannot always be confidently identified to species, there is a risk of inflating the number of species actually collected. For each sample, we reviewed the list of taxa to determine whether they were represented at more then one level of resolution. For example, if an "Unknown Catostomus" was collected, and it was the only representative of the genus at the site, we assigned it as distinct. If any other species of the genus were collected, then we considered the unknown as not distinct. We used only the number of distinct taxa in the sample to calculate any metrics based on species richness.

FISH MULTI-METRIC INDEX DEVELOPMENT

We used a consistent process to develop an FMMIfor each of the three climatic regions. For each candidate metric, we applied the set of predictor variables to a set of reference sites using a random forest model [\(Cutler et al. 2007,](#page-135-10) [Hawkins et al. 2010a\)](#page-135-9). The model provided expected values for the metric (i.e. under least-disturbed conditions) given the particular values of the predictor variables. This approach served to help remove the effects of natural gradients on metric response, which are often confounded with disturbance gradients when expected values for a metric are based solely on a set of regional reference sites [\(Hawkins et al. 2010a\)](#page-135-9). The model for each metric was then applied to the entire set of sites, and the residuals (deviation from predicted) was used as the response value for the metric. We evaluated each metric for its responsiveness to disturbance, i.e., the ability to discern between least-disturbed (reference sites) and more highly disturbed sites [\(following Stoddard et al. 2008\)](#page-136-9). We then selected metrics representing different dimensions of assemblage structure or function to include in the FMMI based on responsiveness and lack of correlation with other metrics, again following Whittier et al. [\(2007\)](#page-137-1) and Stoddard et al. [\(2008\)](#page-136-9).

3.1 Reference Sites for Fish

We modified the base list of reference sites determined for NRSA to eliminate additional fish samples that might not be representative of least disturbed conditions (Table D- 1). The final set of 329 reference sites used for the FMMI development are listed in Appendix D-A.

To validate the random forest model for each metric, we identified a random subset of reference sites (validation sites) within each climatic region and excluded them from model development. We set aside 30 validation sites in the Eastern Highlands, 36 sites in the Plains and Lowlands, and 19 sites in the West region. Applying the model to these reference sites should produce MMI scores that are similar to those sites that wee used to develop the model.

Table D-1. Criteria used to select reference sites for use in developing the FMMI.

3.2 Candidate Metrics

We calculated 162 candidate metrics representing the following dimensions of fish assemblage structure and function [\(following Stoddard et al. 2008\)](#page-136-9):

Species richness Taxonomic composition Habitat guild Trophic guild Reproductive guild Migratory pattern (life history) **Tolerance** Nonnative species

For nearly all metrics, three variants were derived based on all taxa in the sample and for only native taxa in the sample: one based on distinct taxa richness, one based on the percent of individuals in the sample, and one based on the percent of distinct taxa in the sample. For some trophic metrics, additional variants were derived using only taxa that were not considered tolerant to disturbance. We included only those tolerance metrics based on sensitive and tolerant taxa, because the "intermediate tolerance" assignments included taxa with unknown tolerance.

3.3 Predictor variables

A total of 55 predictor variables were initially provided for all NRSA sites (including handpicked sites). These data were provided to us by Dr. Charles Hawkins and his staff at the Western Center for Monitoring and Assessment of Freshwater Ecosystems, Utah State University, Logan, UT. These variables (Appendix E-B) represented the primary natural gradients that are believed to constrain the fish assemblage composition in the absence of human disturbance. The set of predictor variables included those related to watershed area and slope, elevation, latitude and longitude, air temperature, precipitation, relative humidity. There were also model-derived estimates of flow, runoff, and predicted stream temperature. Many variables were estimated at the point level (at the site coordinates) and the watershed level (all values within a particular watershed were aggregated in some fashion). For the FMMI, we constrained the development of the index to only those sites that had both point and watershed level predictors, as we felt that fish might be more responsive to larger-scale natural driver variables than to more site-specific conditions.

In addition to the predictor variables provided, we calculated potential discharge (Q-POTENT_WS) as the product of runoff and watershed area, and included an aggregated ecoregion variable (FW_ECO9; see Figure D 2). We screened the set of predictor variables to eliminate those with large discrepancies in range between the set of reference sites and all other sites, and to eliminate those that had missing values for some sites, as a complete set of predictor variables is required to construct the metric models. The final set of 64 predictor variables that we used to develop the FMMI are identified in boldface type in Appendix D-C.

Figure D-2. Aggregated Omernik ecoregions (Level III) used as predictor variable for NRSA FMMI development and for assigning condition based on MMI score.

3.4 Random Forest Modeling

We used the R statistical package [\(version 2.13.0; R Development Core Team 2011\)](#page-136-10), and the package randomForest [\(version 4.6-6; Liaw and Weiner 2002\)](#page-136-11). We used the set of reference sites (minus those set aside for validation) to develop the metric models. The number of trees to generate was set at 500.

For metrics having a "pseudo-r²" value > 0.10, we applied the metric model to the entire set of sites, and retained the residual values as the modeled metric response value. For metrics with "pseudo r^2 " values \leq 0.10, we retained the original response value.

3.5 Final Metric Selection

For both original and modeled metrics, the mean response values of the set of reference sites and the set of more highly disturbed sites were compared with two-sample *t*-tests (assuming unequal variances). Stoddard et al. [\(2008\)](#page-136-9) present the advantages of using *t* values over other statistics as an indicator of metric responsiveness to disturbance. We also generated a correlation matrix (Pearson *r*) of the metrics to check for redundancy.

To select the final suite of metrics to include in the FMMI, we selected the metric with the largest value of *t* [\(Whittier et al. 2007,](#page-137-1) [Stoddard et al. 2008\)](#page-136-9). The metric with the next largest value of t that was from a different metric class than the previous metric was selected next. If this metric was uncorrelated with the first metric ($r < 0.7$), it was retained in the final set. This process was repeated until there was one metric from each category. If no metric in a class has a *t* value > 3, that category was not represented in the final suite.

Table D-2 presents the final suites of metrics selected for each of the three regional FMMIs. The Plains and Lowlands FMMI does not include a migration strategy metric, while the Eastern Highlands and West FMMIs do not include a composition metric.

We examined the variable importance plots of for each metric in the final set (produced with the R function varImpPlot) to identify the predictor variables that were the most influential in the model. We also examined partial dependence plots of the important predictor variables (produced with the R function partialPlot) to confirm the reasonableness of the relationships between predictor and metric.

Table D-2. Suite of final metrics included in each regional FMMI. Variable names are in parenthesis. Metrics in bold are modeled metrics. Values of *t* are from comparisons of mean values of reference and more highly disturbed sites.

3.6 Metric Scoring

Response values for each of the final suite of metrics was rescaled to a score ranging between 0 and 10. We used the $5th$ and $95th$ percentiles of all sites to set the "floor" (below which a score of 0 was assigned) and "ceiling (above which a score of 10 was assigned) as recommended by Blocksom [\(2003\)](#page-135-11). Response values between the floor and ceiling were assigned a score using linear interpolation.

We summed the metric scores for each site to derive the FMMI score. We then multiplied the FMMI score by (10/number of metrics) to rescale the score to range between 0 and 100 range.

3.7 Sites with Low Fish Abundance

The target population of streams and rivers for NRSA included small headwater streams. Some very small streams may not contain fish even in the absence of human disturbance. We followed the approach described by McCormick et al [\(2001\)](#page-136-12) and used reference sites to estimate a drainage area below which the probability was high that no fish would be present. This approach uses the relationship between a set of four physical habitat variables that characterize habitat volume and the number of fish collected. This relationship defines a habitat volume value below which nearly all sites sampled were devoid of fish. Then this habitat volume value is related to watershed area to determine the drainage area below which streams are expected to be naturally fishless.

Figure D- 3 shows the results of this analysis. The value for the habitat volume index below which almost all sites are fishless is 0.43. When habitat volume is plotted against watershed area, this value corresponds to a watershed area of approximately 2 km². For sites with watershed areas less than 2 km^2 where no fish were collected, we did not report the FMMI score. Otherwise, we assigned an FMMI sore of zero to sites with no fish collected.

Table D-3. Determining the minimum drainage area expected to reliably support the presence of fish (adapted from McCormick et al [\(2001\)](#page-136-12). Variable names are from the NRSA database. Scores for each metric between the upper and lower criteria were estimated by linear interpolation.

Figure D-3. Relationship between small watershed size, reduced habitat volume, and number of fish collected. Fish are not likely to be found in streams with a watershed area of < 2 km2.

4.0 FMMI PERFORMANCE

We evaluated the performance of the regional FMMIs in several ways (Table D-4). Comparing the FMMI scores from set of validation reference sites to those of the set of reference sites used for random forest modeling confirmed that the models were behaving as anticipated. For all three regional FMMIs, the mean values of the validation sites and sites used in modeling were not significantly different based on a two-sample *t*-test assuming unequal variances).

We evaluated the responsiveness of the regional FMMIs by comparing FMMI scores of the set of reference sites to the set of more highly disturbed sites [\(Stoddard et al. 2008\)](#page-136-9). Boxplots (Figure D-4) and two-sample *t* tests (assuming unequal variances) showed that all FMMIs were highly responsive, but the FMMI for the West region was somewhat less responsive than the other two FMMIs (Table D-4).

We estimated precision of the models by calculating the standard deviation of FMMI scores from all reference sites, after standardizing the scores to a mean of 0. The FMMIs all appear to be very precise, with standard deviation value near 0.1 (Table D-4). These values are comparable (or better) than many predictive models of taxa loss [\(Hawkins et al. 2010a\)](#page-135-9).

We evaluated the reproducibility of the regional FMMIs using a set of sites that were visited at least twice during the course of the NRSA project, typically two times in a single year [\(Kaufmann et al.](#page-136-13) [1999,](#page-136-13) [Stoddard et al. 2008\)](#page-136-9). We us a general linear model (PROC GLM, SAS v. 9.12) to obtain estimates of among-site and within-site (from repeat visits) variability. PROC GLM was used because of the highly unbalanced design (only a small subset of sites had repeat visits). We used a nested model (sites within year) where both site and year were random effects. We estimated reproducibility by deriving a "Signal:Noise" (S/N) ratio as (*F* – 1)/*c*, where *F* is the *F*-statistic from the ANOVA, and *c* is a coefficient in the equation used to estimate the expected mean square. If all sites had repeat visits, *c* would equal 2 [\(Kaufmann et al. 1999\)](#page-136-13). If no sites had repeat visits, *c* would equal 1. For the Eastern Highlands, c=1.1326, while for the Plains and Lowlands, *c*=1.0745, and for the West, *c*=1.0753. Values of S/N suggest the regional MMIs are reproducible, with values between 4 (West) and 8 (Plains and lowlands; Table E-4).

Performance Characteristic	Eastern	Plains and	West
	Highlands	Lowlands	FMMI
	FMMI	FMMI	
Validation reference sites vs. reference sites used	$t=0.13$	$t = 1.19$	$t = -0.14$
in metric modeling			
Reference sites vs. more highly disturbed sites	$t = 18.2$	$t = 17.2$	$t = 11.3$
Model precision	0.091	0.120	0.075
Reproducibility (Signal: Noise)	5.2	8.0	4.1

Table D-4. Performance statistics for the three regional FMMIs.

Figure D-4. Boxplots comparing FMMI scores of reference sites to more highly disturbed sites.

We felt it important to examine the performance of the component metrics across the range of stream sizes sampled for NRSA. The potential exists for bias in the FMMI due to different fish species pools being available for larger rivers versus smaller streams, Differences across the size range might also result from the different sampling protocols that were used (wadeable, large wadeable, and boatable). We used the set of reference sites to examine patters in metric response values across Strahler order categories. The distribution of metric response values and FMMI scores among stream order classes does not indicate a bias due to either stream size or sampling method (Figures D- 5-7).

We looked at the distribution of FMMI scores across Strahler order categories as well. We did not expect to see much similarity, as disturbance is typically confounded with stream size. This pattern is evident in the Eastern Highlands and the West (Figure D-8).

Figure D-5. Component metrics of the Eastern Highlands FMMI versus Strahler Order category,

Figure D-6. Component metrics of the Plains and Lowlands FMMI versus Strahler Order category,

Figure D-7. Component metrics of the West region FMMI versus Strahler Order category,

Figure D- 8. FMMI scores versus Strahler order category.

THRESHOLDS FOR ASSIGNING ECOLOGICAL CONDITION

For NRSA, ecological condition is based on the deviation from least disturbed condition [\(Stoddard et](#page-136-14) [al. 2006,](#page-136-14) [Hawkins et al. 2010b\)](#page-135-12). Within each of the three climatic regions, thresholds for defining "Good" condition (similar to least-disturbed) and "Poor" condition (substantially different from least disturbed) are based on the distribution of FMMI scores in reference sites in each of the nine aggregated ecoregions (Figure D-2). The threshold for "good" condition is equal to the $25th$ percentile of the distribution of reference sites withn an aggregated ecoregion. The threshold for "poor" condition is set as being below the $5th$ percentile of the distribution of reference sites within an aggregated ecoregion.

Table D-5 presents the threshold values for the regional FMMIs. Sites with scores between the two threshold values were assigned a condition class of "Fair" (indeterminate). In general, threshold values within a climatic regions differ from each other to prevent combining aggregated ecoregions to reduce the number of different thresholds being used. Three aggregated ecoregion contain < 30 reference sites, so the threshold values (particularly those for Fair/Poor) are based on very few sites. In three regions, the difference between the two thresholds is < 5 (Xeric=2.9). Table D-5. Thresholds for assigning ecological condition based on the distribution of regional FMMI scores in reference sites. Aggregated ecoregions are shown in

Aggregated Ecoregion	Good/Fair (25 th percentile)	Fair/Poor (5 th percentile)
Eastern Highlands		
Northern Appalachians (60)	65.4	60.2
Southern Appalachian (94)	66.1	55.4
Plains and Lowlands		
Coastal Plains (39)	55.3	46.9
Northern Plains (42)	54.6	48.3
Southern Plains (43)	54.2	49.2
Temperate Plains (28)	60.3	52

Figures D1 and D-2. Sample sizes are in parentheses)

5.0 LITERATURE CITED

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. EPA 841/B-99/002, Office of Water. US Environmental Protection Agency, Washington, DC.

Blocksom, K. A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. Environmental Management 31:0670-0682.

Bramblett, R. G., T. R. Johnson, A. V. Zale, and D. G. Heggem. 2005. Development and evaluation of a fish assemblage index of biotic integrity for northwestern Great Plains streams. Transactions of the American Fisheries Society 134:624-640.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

Emery, E. B., T. P. Simon, F. H. McCormick, P. L. Angermeier, J. E. DeShon, C. O. Yoder, R. E. Sanders, W. D. Pearson, G. D. Hickman, R. J. Reash, and J. A. Thomas. 2003. Development of a multimetric index for assessing the biological condition of the Ohio River. Transactions of the American Fisheries Society 132:791-808.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. Ecological Applications 16:1277-1294.

Hawkins, C. P., Y. Cao, and B. Roper. 2010a. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066-1085.

Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010b. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29:312-343.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:21-27.

Karr, J. R. 1991. Biological integrity: a long neglected aspect of water resource management. Ecological Applications 1:66-84.

Karr, J. R. 1999. Defining and measuring river health. Freshwater Biology 41:221-234.

Karr, J. R. and E. W. Chu. 2000. Sustaining living rivers. Hydrobiologia 422/423:1-14.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Special Publication 5, Illinois Natural History Survey, Champaign, Illinois.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA 620/R-99/003, Office of Research and Development, US Environmental Protection Agency, Washington, DC.

Liaw, A. and M. Weiner. 2002. Classification and Regression by randomForest. R News 2:18-22.

Lyons, J., R. R. Piette, and K. W. Niermeyer. 2001. Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large warmwater rivers. Transactions of the American Fisheries Society 130:1077-1094.

McCormick, F. H., R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, and A. T. Herlihy. 2001. Development of an index of biotic integrity for the Mid-Atlantic Highlands region. Transactions of the American Fisheries Society 130:857-877.

Meador, M. R. and D. M. Carlisle. 2009. Predictive Models for Fish Assemblages in Eastern U.S. Streams: Implications for Assessing Biodiversity. Transactions of the American Fisheries Society 138:725-740.

Mebane, C. A., T. R. Maret, and R. M. Hughes. 2003. An Index of Biological Integrity (IBI) for Pacific Northwest Rivers. Transactions of the American Fisheries Society 132:239-261.

Nelson, J. S., E. J. Crossman, H. Espinosa-Pérez, L. T. Findley, C. R. Gilbert, R. K. Lea, and J. D. Williams. 2004. Common and Scientific Names of Fishes from the United States Canada and Mexico. Sixth edition. Special Publication 29, American Fisheries Society, Bethesda, Maryland.

Page, L. M. and B. M. Burr. 1991. A field guide to freshwater fishes of North America north of Mexico. Houghton Mifflin, Boston, Massachusetts.

Pearson, M. S., T. R. Angradi, D. W. Bolgrien, T. M. Jicha, D. L. Taylor, M. F. Moffett, and B. H. Hill. 2011. Multimetric Fish Indices for Midcontinent (USA) Great Rivers. Transactions of the American Fisheries Society 140:1547-1564.

Pont, D., R. M. Hughes, T. R. Whittier, and S. Schmutz. 2009. A Predictive Index of Biotic Integrity Model for Aquatic-Vertebrate Assemblages of Western U.S. Streams. Transactions of the American Fisheries Society 138:292-305.

Pont, D., B. Hugueny, and C. Rogers. 2007. Development of a fish-based index for the assessment of river health in Europe: the European Fish Index. Fisheries Management and Ecology 14:427-439. R Development Core Team. 2011. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL [http://www.R](http://www.r-project.org/)[project.org.](http://www.r-project.org/)

Roset, N., G. Grenouillet, D. Goffaux, D. Pont, and P. Kestemont. 2007. A review of existing fish assemblage indicators and methodologies. Fisheries Management and Ecology 14:393-405. Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878-891.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16:1267-1276.

USEPA (United States Environmental Protection Agency). 2009. National Rivers and Streams Assessment: Field Operations Manual. EPA 841/B-04/004, Office of Water and Office of Environmental Information, US Environmental Protection Agency, Washington, DC.

Whittier, T. R., R. M. Hughes, J. L. Stoddard, G. A. Lomnicky, D. V. Peck, and A. T. Herlihy. 2007. A structured approach for developing indices of biotic integrity--three examples from western streams and rivers in the USA. Transactions of the American Fisheries Society 136:718-735.

APPENDIX D4 – NRSAMACROINVERTEBRATE COMMUNITY ASSEMBLAGE

Background Information

The taxonomic composition and relative abundance of different taxa that comprise the benthic macroinvertebrate assemblage present in a stream have been used extensively in North America, Europe, and Australia to assess how human activities affect ecological condition (Barbour et al. 1995, 1999; Karr and Chu 1999). Two principal types of ecological assessment tools to assess condition based on benthic macroinvertebrates are currently prevalent: multimetric indices and predictive models of taxa richness. The purpose of these indicators is to present the complex community taxonomic data represented within an assemblage in a way that is understandable and informative to resource managers and the public. The following sections provide a general overview of the approaches used to develop ecological indicators based on benthic macroinvertebrate assemblages, followed by details regarding data preparation and the process used for each approach to arrive at a final indicator.

1.1 Overview of the IBI and O/E Predictive Model Approaches

Multimetric indicators have been used in the U.S. to assess condition based on fish and macroinvertebrate assemblage data (e.g., Karr and Chu, 1999; Barbour et al., 1999; Barbour et al., 1995). The multimetric approach involves summarizing various assemblage attributes (e.g., composition, tolerance to disturbance, trophic and habitat preferences) as individual "metrics" or measures of the biological community. Candidate metrics are then evaluated for various aspects of performance and a subset of the best performing metrics are then combined into an index, referred to as a multimetric index or MMI. For NRSA, the MMI developed in the WSA was used to generate the population estimates used in the assessment. The WSA MMI is detailed in Stoddard et al. (2008).

The predictive model approach was initially developed in Europe and Australia, and is becoming more prevalent within the U.S. The approach estimates the expected taxonomic composition of an assemblage in the absence of human stressors (Hawkins et al., 2000; Wright, 2000), using a set of "least disturbed" sites and other variables related natural gradients (such as elevation, stream size, stream gradient, latitude, longitude). The resulting models are then used to estimate the expected taxa composition (expressed as taxa richness) at each stream site sampled. The number of expected taxa actually observed at a site is compared to the total number of expected taxa as an Observed:Expected ratio (O/E index). Departures from a ratio of 1.0 indicate that the taxonomic composition in a stream sample differs from that expected under less disturbed conditions.

2.0 Data Preparation

2.1 Standardizing Counts

The number of individuals in a sample was standardized to a constant number to provide an adequate number of individuals that was the same for the most samples and that could be used for both multimetric index development and O/E predictive modeling index. A subsampling technique involving random sampling without replacement was used to extract a true "fixed count" of 300 individuals from the total number of individuals enumerated for a sample (target lab count was 500 individuals). Samples that did not contain at least 300 individuals were used in the assessment because low counts can indicate a response to one or more stressors. Only those sites with at least 250 individuals, however, were used as reference sites.

2.2 Operational Taxonomic Units

For the predictive model approach, it was necessary to combine taxa to a coarser level of common taxonomy. This new combination of taxa is termed an "operational taxonomic unit" or OTU, and results in fewer taxa than are present in the initial benthic macroinvertebrate count data.

2.3 Autecological Characteristics

Autecological characteristics refer to specific ecological requirements or preferences of a taxon for habitat preference, feeding behavior, and tolerance to human disturbance. These characteristics are prerequisites for identifying and calculating many metrics. A number of state/ regional organizations and research centers have developed autecological characteristics for benthic macroinvertebrates in their region. For the WSA and NRSA, a consistent "national" list of characteristics that consolidated and reconciled any discrepancies among the regional lists was needed before developing and calibrating certain biological metrics and constructing an MMI.

Members of the data analysis group pulled together autecological information from five existing sources: the USEPA Rapid Bioassessment Protocols document, the National Ambient Water Quality Assessment (NAWQA) national and northwest lists, the Utah State University list, and the EMAP Mid-Atlantic Highlands (MAHA) and Mid-Atlantic Integrated Assessment (MAIA) list. These five were chosen because they were thought to be the most independent of each other and the most inclusive. A single national-level list was developed based on the following decision rules:

2.3.1 Tolerance Values

Tolerance value assignments followed the convention for macroinvertebrates, ranging between 0 (least tolerant or most sensitive) to 10 (most tolerant). For each taxon, tolerance values from all five sources were reviewed and a final assignment made according to the following rules:

If values from different lists were all $<$ 3 (sensitive), final value = mean;

If values from different lists were all >3 and <7 (facultative), final value = mean;

If values from different lists were all >7 (tolerant), final value = mean;

If values from different lists spanned sensitive, facultative, and tolerant categories, BPJ was used, along with alternative sources of information (if available) to assign a final tolerance value. .

Tolerance values of 0 to ≤3 were considered "sensitive" or "intolerant". Tolerance values ≥7 to 10 were considered "tolerant," and values in between were considered "facultative."

2.3.2 Functional Feeding Group and Habitat Preferences

In many cases, there was agreement among the five data sources. When discrepancies in functional feeding group (FFG) or habitat preference ("habit") assignments among the five primary data sources were identified, a final assignment was made based on the most prevalent assignment. . In cases where there was no prevalent assignment, the workgroup examined why disagreements existed; flagged the taxon, and used best professional judgment to make the final assignment.

3.0 MultiMetric Index Development

3.1 Regional MultiMetric Development

The autecology and taxonomic resolution used in WSA was applied to the NRSA macronvertebrate 300 fixed count data to calculate the community metrics used to calculate the MMI. In the WSA, a best ecoregion MMI was developed by summing the 6 metrics that performed best in that ecoregion (the national aggregate 9 ecoregions). Each metrics was scored on a 0-10 scale and then summed and normalized to a 0-100 scale to calculate the final MMI. Details of this process are described in Stoddard et al. (2008). The final metrics used in each ecoregion are summarized in Table C-1. The NRSA MMI was calculated in the same way as the WSA MMI. Based on NRSA revisit data, the MMI had a S:N ratio of 2.8 and a pooled standard deviation of 10.0 (out of 0-100).

Table C-1. Six benthic community metrics used in for the NRSA and WSA MMI in each of the nine aggregate ecoregions

3.2 Modeling of MMI condition class thresholds for the Wadeable Streams Assessment Previous large-scale assessments have converted MMI scores into classes of assemblage condition by comparing those scores to the distribution of scores observed at least-disturbed reference sites. If a site's MMI score was less then the $5th$ percentile of the reference distribution, it was classified as 'poor' condition, between the 5th and 25th percentile was classified as 'fair', and greater than or equal to the 25th percentile was classified as 'good'. This approach assumes that the distribution of MMI scores at reference sites reflects an approximately equal, minimum level of human disturbance across those sites. But this assumption did not appear to be valid for some of the 9 WSA regions, which was confirmed by state and regional parties at meetings to review the draft results.

For the WSA, the project team performed a principal components analysis (PCA) of 9 habitat and water chemistry variables that had originally been used to select IBI reference sites. The first principal component (Factor 1) of this PCA represented a generalized gradient of human disturbance. MMI scores at the reference sites were weakly, but significantly, related to this disturbance gradient in 5 of the 9 aggregate regions. Thus, MMI reference distributions from these regions are biased downward, because they include somewhat disturbed sites which have lower MMI scores. As part of the WSA, Herlihy et al. (2008) developed a process that used the PCA disturbance gradient scores to reduce the effects of disturbance within the reference site population. The process used multiple regression modeling to develop adjusted thresholds analogous to the $5th$ and 25th percentiles of reference sites in each ecoregion. These adjusted thresholds were used in WSA and the same thresholds were used in NRSA to define good, fair and poor condition based on the benthic MMI. The process for calculating these adjusted thresholds is detailed in Herlihy et al. (2008) and the threshold values for each ecoregion used in WSA and the NRSA report are given in Table C- 2.

Table C- 2. Threshold values for the 9 regional benthic MMIs. Any site with an MMI score that was not good or poor was considered "Fair".

4.0 O/E: Predictive (RIVPACS) Models

In addition to the benthic macroinvertebrate MMI approach, predictive O/E modeling was used to assess benthic macroinvertebrate condition for the NRSA. The O/E model compares the observed benthic assemblage at a site to an expected assemblage derived from a population of reference sites. Stressors and anthropogenic impacts lead to a reduction in the number of taxa that are expected to be present under reference conditions. The predictive model approach is used by several states and is a primary assessment tool of Great Britain and Australia.

The O/E ratio predicted by the model for any site expresses the number of taxa found at that site (O), as a proportion of the number that would be expected (E) if the site was in least-disturbed condition. Ideally, a site in reference condition has $O/E = 1.0$. An O/E value of 0.70 indicates that 70% of the "expected" taxa at a site were actually observed at the site. This is interpreted as a 30% loss of taxa relative to the site's predicted reference condition. However, O/E values vary among reference sites themselves, around the idealized value of 1.0, because such sites rarely conform to an idealized reference condition, and because of model error and sampling variation. The standard deviation of O/E (Table C-3) indicates the breadth of O/E variation at reference sites. Thus, the O/E value of an individual site should not be interpreted as (1 – taxa loss) without taking account of this variability in O/E. Individual O/E values are most reliably interpreted relative to the entire O/E distribution for reference sites.

A nationally-distributed collection of reference sites was first identified, drawn from a pool of sites whose macroinvertebrates were sampled using EMAP protocols. This pool included only NRSA, WSA, EMAP-West, STAR-Hawkins, USGS NAWQA, and MAHA/MAIA sites. One hundred reference sites were set aside to validate the models, and the remaining reference sites were used to calibrate the models (Table C-3). Each site contributed a single sampled macroinvertebrate assemblage to model calibration and validation. Each sampled macroinvertebrate assemblage comprising more than 300 identified individuals was randomly subsampled to yield 300 individuals. 300-count subsamples were used to build models and assess all NRSA sites.

The predictive modeling approach assumes that expected assemblages vary across reference sites throughout a region, due to natural (non-anthropogenic) environmental features such as geology, soil type, elevation and precipitation. To model these effects, the approach first classifies reference sites based on similarities of their macroinvertebrate assemblages (Table C-3). A random forest model is then built to predict the membership of any site in these classes, using natural environmental features as predictor variables (Table C-3). The predicted occurrence probability of a reference taxon at a site is then predicted to be the weighted average of that taxon's occurrence frequencies in all reference site classes, using the site's predicted group membership probabilities in the classes as weights. Finally, E for any site is the sum, over a subset of reference taxa, of predicted taxon occurrence probabilities. O is the number of taxa in that subset that were observed to be present at the site. The subset of reference taxa used for any site was defined as those taxa with predicted occurrence probabilities exceeding 0.5 at that site.

Final predictive models performed better than corresponding null models (no adjustment for naturalfactor effects), as judged by their smaller standard deviation of O/E across calibration sites (Table C-3).

Similar to the IBI, two scaled approaches were used to develop the O/E model. A national model was initially developed to predict taxa loss at sites. Three models were developed for NRSA usage, together covering the contiguous USA (Table C-3). The regional models performed better, and were used in the NRSA to predict taxa loss at the sites.

Table C-3. NRSA predictive models.

5.0 Literature Cited

Herlihy, A.T., S.G. Paulsen, J. Van Sickle, J.L. Stoddard, C.P. Hawkins, and L.L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference condition approach at a continental scale. Journal of the North American Benthological Society 27:860-877.

Stoddard, J.L., A.T. Herlihy, D.V. Peck, R.M. Hughes, T.R. Whittier, and E. Tarquinio. 2008. A process for creating multi-metric indices for large scale aquatic surveys. Journal of the North American Benthological Society 27:878-891.

L.L. Yuan, C.P. Hawkins, J. Van Sickle. 2008. Effects of regionalization decisions on an O/E index for the US national assessment. Journal of the North American Benthological Society 27:892-905.
APPENDIX D5 – NRSAWATER CHEMISTRY ANALYSIS

Background Information

Four chemical stressors are summarized in the NRSA report; total nitrogen, total phosphorus, acidity and salinity. Criteria values and class definitions were identical to those used in the Wadeable Streams Assessment (WSA) as described below. Threshold were established for each of the nine ecoregions reported on in the NRSA and WSA.

Threshold Development

2.1 Acidity and Salinity

For acidity, criteria values were determined based on values derived during the NAPAP program. Sites with acid neutralizing capacity (ANC) less than zero were considered acidic. Acidic sites with dissolved organic carbon (DOC) greater than 10 mg/L were classified as organically acidic (natural). Acidic sites with DOC less than 10 and sulfate less than 300 µeq/L were classified as acidic deposition impacted, while those with sulfate above 300 µeq/L were considered acid mine drainage impacted. Sites with ANC between 0 and 25 µeq/L and DOC less than 10 mg/L were considered acidic deposition influenced but not currently acidic. These low ANC sites typically become acidic during high flow events (episodic acidity).

Salinity and nutrient classes were divided into either good, fair, or poor classes. Salinity classes were defined by specific conductance using ecoregion specific values (Table F-1).

2.2 Total Nitrogen and Phosphorus

Total nitrogen and phosphorus were classified using a method similar to that used for macroinvertebrate IBI classes using deviation from reference by aggregate ecoregion. For nutrients, the value at the 25th percentile of the reference distribution was selected for each region to define the least disturbed condition class (good-fair boundary). The 5th percentile of the reference distribution defines the most disturbed condition class (Table F-1). For setting nutrient class boundaries, only reference sites from the screened WSA dataset were used. Since nutrients were the focus, the two nutrient criteria used in defining reference sites were dropped and the other seven criteria used by themselves to identify a set of "nutrient reference sites." Before calculating percentiles from this set of sites, outliers (values outside 1.5 times the interquartile range) were removed. Percentiles were calculated in WSA from WSA data. They were not updated or changed based on NRSA data.

Ecoregion	Salinity as Conductivity $($ uS/cm $)$ Good-Fair	- Salinity as Conductivity (uS/cm) Fair-Poor	N Total (ug/L) Good- Fair	N Total (ug/L) Fair-Poor	P Total (ug/L) Good- Fair	P Total (ug/L) Fair-Poor
CPL	500	1000	1092	2078	56.3	108
NAP	500	1000	329	441	8.2	15.7
SAP	500	1000	296	535	17.8	24.4
UMW	500	1000	716	1300	21.6	44.7
TPL	1000	2000	1750	3210	165	338
NPL	1000	2000	948	1570	91.8	183
SPL	1000	2000	698	1570	52.0	95.0
WMT	500	1000	131	229	14.0	36.0
XER	500	1000	246	462	35.5	70.0

Table F-1. Nutrient and Salinity Category Criteria for NRSA Assessment

Literature Cited

A.T. Herlihy, J.C. Sifneos. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. Journal of the North American Benthological Society. Vol. 27, Issue 4.