Assessing stream restoration and the influence of scale, variable choice, and comparison sites

KRISTOPHER K. WRIGHT

Department of Biology, University of Wisconsin-Platteville, 1 University Plaza, Platteville, Wisconsin 53818 USA

Citation: Wright, K. K. 2021. Assessing stream restoration and the influence of scale, variable choice, and comparison sites. Ecosphere 12(3):e03440. 10.1002/ecs2.3440

Abstract. Assessing stream restoration remains an on-going challenge. Conclusions about the effectiveness of stream restoration efforts vary throughout the literature. This uncertainty is likely due, in part, to the spectrum of scales at which restorations have been assessed, the different variables used for assessment, and the varying types of sites used for comparisons. Mitigating such uncertainty with stream restoration assessment requires perspectives that consider multiple sites, variables, and scales using approaches that are transferable and comparable. From 2004 to 2018, surveys of stream habitat and fish at 44 restored and unrestored sites were conducted in two adjacent watersheds in southwestern Wisconsin, USA. Using non-parametric randomization tests, means of 42 variables were compared between (1) watershed vs. watershed sites; (2) all restored sites vs. all unrestored sites, (3) restored sites vs. pre-treatment surveys, (4) restored sites vs. unrestored control sites, and (5) pre-treatment surveys vs. unrestored control sites. When applicable, these comparisons were also made at nested scales ranging from all sites in both watersheds to only contiguous sites with common survey periods. Comparing sites between the two watersheds revealed natural variability between watersheds with similar geology and land-use practices, though unexpectedly more differences were found between surveys of restored sites from the two watersheds than between unrestored sites. The most differences between surveys of restored and unrestored sites were observed at larger scales of consideration, but within a given watershed. Although relatively few variables differed between pre-treatment surveys and control sites, there were marked differences when each type of unrestored sites was separately compared with restored sites. Habitat variable differences were more robust and consistent than fish community variables across all comparisons. Deviating from traditional multivariate approaches and independently comparing a broad spectrum of individual variable means at various, nested scales underscored the importance of context and variable choice when assessing stream restoration.

Key words: assessment; Driftless Area; fish; habitat; monitoring; randomization tests; river restoration; scale; stream restoration; streams; variables.

Received 23 April 2020; accepted 25 November 2020; final version received 18 January 2021. Corresponding Editor: Wyatt F. Cross.

Copyright: © 2021 The Author. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† E-mail: wrightk@uwplatt.edu

INTRODUCTION

While stream restoration has become a global, multibillion-dollar enterprise, the evidence of how restoration impacts streams and their communities is highly variable (Roni 2019). There are many reasons behind this uncertainty (Kail et al. 2015, Friberg et al. 2016, Shirey et al. 2016, Höckendorf et al. 2017, Rubin et al. 2017). Different geology and geography provide unique foundational conditions among streams and watersheds.
Land-use impacts that degrade streams vary in type, intensity, spatial extent, and time. Restoration projects differ depending on goals, scales, resources, access, and practitioners. Historic degradation may compromise regional species pools (Sundermann et al. 2011, Stoll et al. 2013, 2014) and connectivity (Tonkin et al. 2014). Finally, assessment strategies and outcomes may be non-existent, limited in scope, unreported, or based on analyses and measures of success that vary among projects (Roni et al. 2005, Jähnig et al. 2011, Rubin et al. 2017).

Progress has been made addressing these issues over the past 20 yr. Literature on stream restoration has grown considerably (e.g., Google Scholar search results for “river restoration” are two times greater from 2017 to 2018 than from the entire 20-year span 1980–1999). Efforts to conduct and report long-term pre- and post-evaluations of stream restoration projects have begun to fill a critical void (e.g., Shirey et al. 2016, Hockendorf et al. 2017). Meta-analyses of past restoration projects have shed light on outcomes, challenges, inconsistencies, and future directions (e.g., Simaika et al. 2015, Thomas et al. 2015, Rubin et al. 2017, Manfrin et al. 2019). Terminology and guidelines have been proposed that attempt to both embrace stream variability and provide common contexts for measuring restoration success (e.g., Palmer et al. 2005, Roni et al. 2005, Weber et al. 2018, Rydgren et al. 2019).

Despite such progress, there remains a need for more published data that address lingering, synergistic questions regarding assessment of stream restoration. How does the inherent variability of streams influence restoration assessment? At what frequency and duration should sites be assessed before and after restoration? Which variables should be measured to provide meaningful and comparable assessments? Should restored areas be compared to undisturbed reference areas, pre-treatment baseline conditions, or degraded control sites? How should restoration assessment data be analyzed and presented? How do changes in both spatial and temporal scales impact interpretations of stream restoration assessments? Answering these questions will require contributions from restoration assessments that measure multiple variables (physical, chemical, and/or biological) from various sites (restored and unrestored) over time.

This study examined restoration and monitoring efforts in two adjacent, small (<1600 km²) watersheds of the Upper Mississippi River Basin in Southwest Wisconsin, USA that began in the early 2000s. As with most restoration efforts around the world, individual restorations in these watersheds occurred at the local (or reach) scale and targeted erosion, floodplain connectivity, water quality, and in-stream habitat for aquatic organisms, particularly trout (Mauldin and Hastings 2014; Wisconsin Department of Natural Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/). Since 2004, stream habitat and fish were surveyed at restored and unrestored sites within the watersheds. This monitoring ranged from single surveys at individual sites to multiple surveys over time within a nearly 10-km contiguous section of stream.

With some rare exceptions (e.g., Hasselquist et al. 2015, Manfrin et al. 2019), published stream restoration assessments typically do not provide nested-scale perspectives. Individual assessments are necessarily constrained in space and/or time, while meta-analyses synthesize outcomes from various regions, watersheds, and/or sources (Roni et al. 2005). Consequently, little is known about how different scales of consideration impact interpretations of stream restoration assessments. Data from this study, combined with a statistical approach rarely used in stream ecology, enabled a unique examination of how different measured variables compared between restored and unrestored sites over multiple scales of consideration. It was hypothesized that at larger scales, fewer measured variables would differ between restored and unrestored sites given greater variability (sensu Wiens 1989, Roni et al. 2005, Ruegg et al. 2016). Presumably as landscape and among-site variability decreases within smaller scales, the impacts of restoration on measured variables should be more apparent between restored and unrestored sites.

Which variables allow for meaningful and comparable assessments is a topic that has received much attention (e.g., Roni et al. 2005, Wohl et al. 2015, Friberg et al. 2016, Rubin et al. 2017, Manfrin et al. 2019). Despite a lack of consensus on a specific set of variables, it seems that habitat characteristics and fish traits may be more robust measures of restoration impacts than the diversity or abundance of macroinvertebrate and
fish communities (e.g., Jähnig et al. 2011, Friberg et al. 2016, Manfrin et al. 2019). This study tested aspects of this notion by using numerous, commonly measured habitat characteristics and the abundance of multiple, regionally ubiquitous fish species for comparisons across various scales.

A review by Rubin et al. (2017) of 26 stream restoration studies highlighted the importance of acknowledging natural variability and providing appropriate comparisons to restored areas. To provide further insight on how spatial variability and choice of unrestored sites influence interpretations of restoration assessments, this study conducted additional comparisons between watersheds and between different types of unrestored sites. Comparing surveys of disturbed, unrestored sites between two adjacent watersheds was intended to reveal underlying, natural variability that may exist between sites from two relatively small watersheds with similar geology and land-use practices. It was then anticipated there would be less variability comparing surveys of restored sites from the two watersheds given their similar restoration objectives and approaches. Lastly, assessments of restoration are typically predicated on comparisons of restored conditions to unrestored conditions; however, Rubin et al. (2017) found assessments varied widely in the use of undisturbed reference areas, pre-treatment surveys, or unrestored control sites as representatives of unrestored conditions. To examine how choice of unrestored conditions may influence restoration assessment, this study compared surveys of restored sites to both pre-treatment surveys and control sites in combination and independently. Additionally, pre-treatment surveys were directly compared with control sites. It was expected that there would be minimal differences between the two different types of unrestored sites, and consequently, they were anticipated to perform similarly when compared independently with surveys from restored sites.

**Methods**

**Study area**

Since European settlement in the early 1800s, land-use practices (i.e., agriculture, deforestation, and mining) dramatically impacted watersheds in the Driftless Area of the Upper Mississippi River Basin. The unique geology and topography of the Driftless Area likely exacerbated the magnitude of these impacts (Juckem et al. 2008). The region avoided the most recent glacial activity resulting in groundwater-fed watersheds with steep headwaters, narrow valley floodplains, and rolling hills with easily erodible, silty soils. Consequently, conversion of prairies and oak savannas to agricultural land contributed to major changes in stream structure and function (Trimble 1993, Benedetti 1993, Knox 2001). Throughout the 1900s, federal and state funded programs encouraged controlling erosion and flooding, and enhancing adult trout habitat (Thorn et al. 1997, Juckem et al. 2008; sensu Vetranio 1988, Hunt 1993). However, limited public land and reluctant private landowners restricted the extent of these efforts (Thorn et al. 1997, Lyons et al. 2000). This legacy provided the context for trout stream management in the Driftless Area of the 21st century.

The 506-km² Blue River watershed of Southwest Wisconsin, USA (43°02’ N, 90°28’ W) drains portions of Grant and Iowa Counties into the Wisconsin River (Fig. 1). Soils are silty and sandy loams over sandstone and dolomite. Land cover is roughly 40% forest, 40% agriculture (crops and cattle), 14% grasslands, and 4% wetlands (Wisconsin Department of Natural Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/). With a population around 10,000, there is very little development in the watershed. The rural setting coupled with nearly 97 km of trout streams has made the Blue River a popular destination for recreational anglers throughout the Midwest (Wisconsin Department of Natural Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/). Directly to the south and west of the Blue River watershed are the Platte River and Grant River watersheds (42°47’ N, 90°32’ W and 42°47’ N, 90°50’ W, respectively). These two adjacent basins are often considered together (Grant-Platte watershed) given their proximity within Grant County and terminal inputs to the Mississippi River within 10 km of each other (Fig. 1). The 1560-km² combined Grant-Platte watershed has predominantly silty loam soils, though occasionally at shallow depths over bedrock and clay subsoils. Agricultural land cover is over 70% in large portions of the watershed, with areas of deciduous forest,
grassland, and oak savanna scattered throughout (Wisconsin Department of Natural Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/). Lotic habitat diversity within the watershed is reflected by over 110 km of trout streams and over 600 km of warm water fisheries (Wisconsin Department of Natural Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/). Although trout streams occur in both the Blue and Grant-Platte watersheds, stream bank erosion, agricultural runoff, overgrazing of streambanks, and sediment loading from croplands potentially compromise many of those streams (Wisconsin Department of Natural Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/).
Resources, Watershed basins, https://dnr.wi.gov/topic/Watersheds/basins/).

To mitigate these impacts, the Harry and Laura Nohr Chapter of Trout Unlimited (HLN Chapter) cooperated with willing landowners, numerous other Trout Unlimited chapters, Wisconsin Department of Natural Resources, Natural Resources Conservation Service, and others to implement a multi-year, stream restoration plan in the early 2000s. The HLN Chapter restoration plan emphasized reducing bank erosion and sedimentation, improving water quality, and enhancing in-stream habitat for both native brook trout (Salvelinus fontinalis) and non-native brown trout (Salmo trutta). Restorations were implemented at the reach scale (delineated by property ownership), varying in length from approximately 500 m to 2 km. All restorations involved tapering banks, thinning or removing riparian trees and bushes, stabilizing meander bends, installing under-bank structures, and adding rock or log structures for habitat and/or flow deflection. Some restorations also included backwater habitats, floodplain scrapes, hibernacula, and/or cattle crossings. To date, the HLN Chapter has restored nearly 23 km of streams (through some 20 properties) in the Blue River and Grant-Platte watersheds.

**Sites and sampling**

All data from 2004 to 2018 were collected during summer-flow conditions (June, July, and/or August) at 300 m (stream length) sites (Figs. 2, 3; Appendix S1: Tables S1, S2). Each of the 44 wadable, 1st–3rd stream order, study sites represented an individual landowner’s property. Sites were located toward the downstream end of each property so that observations more likely represented proximate conditions than the upstream property. Six streams had sampling sites located on adjacent, contiguous properties: Blue River, Rountree Branch, Sixmile Branch, McPherson Branch, Borah Creek, and Bronson Creek (Appendix S1: Tables S1, S2). The remaining sites were either on non-contiguous properties or were single sites from individual streams within the study area.

Habitat measurements at each study site occurred at 12 evenly spaced transects perpendicular to the stream and followed slightly modified procedures of Wisconsin Department of Natural Resources (2002). Wetted channel width, percent macrophyte cover, and percent overhead vegetation cover were single measurements (width) or visual estimates (macrophyte and overhead cover) spanning each transect. Bank erosion was the vertical height of exposed soil on both banks at each transect. Depth, embeddedness, and substrate composition were measured at four equally spaced points across each transect. Embeddedness was the depth into the streambed that could be penetrated with a meter stick. Substrate composition was the estimated proportions of bedrock, boulder, large rock, cobble, gravel, sand, and fines at each point; classifications were based on a modified Wentworth scale (Wentworth 1922). Width to depth ratio (width:depth) was determined by dividing a site’s average wetted width by average depth. Approximate site volume (m$^3$) was the average wetted width $\times$ average depth $\times$ site length.

Fish were surveyed by single pass, backpack electro-fishing through each 300-m site. One to three electro-fishers were used depending on site width. All fish were identified to species and counted. All trout $<100$ mm total length were collectively identified as young-of-the-year (YOY) trout and counted separately from those longer than 100 mm. Species counts were standardized across the various sites by determining the density (no./m$^3$) based on the approximated volume of each site for each sample period. Relative abundance (%) was also calculated for brown trout $>100$ mm, YOY trout, white suckers (Catostomus commersonii), creek chubs (Semotilus atromaculatus), common shiners (Luxilus cornutus), and mottled sculpin (Cottus bathidii) at each site for each sample period.

Sites from each sample period were identified as restored or unrestored (Appendix S1: Tables S1, S2). Sites were considered restored only after restoration activities had been completed throughout the entire property; whereas, unrestored sites were either those sampled prior to restoration activity (hereafter “pre-treatment surveys”) or alternative, degraded sites remaining unrestored beyond 2020 (hereafter “control sites”). Twenty-nine site locations were unrestored throughout the duration of the study; 15 of those were where pre-treatment surveys occurred and 14 were control sites (Appendix S1: Tables S1, S2). Across all sample periods
(including sampling the same site more than once per year), there were 122 total different sample units (surveys) with 58 unrestored and 64 restored (Appendix S1: Tables S1, S2).

Although not all sites were sampled each year, multiple sites were sampled numerous times over the study period (Appendix S1: Tables S1, S2). Notably, Blue River restored sites BR2 and
BR3 have each been surveyed at least 10 times over 14 yr, and control site BR 4 was surveyed seven times over a 10-year span. In 19 cases across all sites, surveys were done twice within a given year at the same site (Appendix S1: Tables S1, S2).

When feasible, stream habitat and fish were surveyed separately within a week for the same sample period. Fish surveys occurred most consistently and were only missed on four occasions: 2007 site RB10; August 2016 site BR9; 2017 site BR5; and 2018 site BR13. Habitat surveys were

Fig. 3. Relative locations of study sites within the Grant-Platte watershed and their restoration status. Not drawn to scale.
not conducted on 13 occasions: June 2011 sites BR1, BR8, BR 9; 2012 site BR10; 2017 sites BR11, BRT1, August BO2; and 2018 sites BR2, BR5, MP2, SX2, August BO2 and BO3.

Data analysis

Although multivariate analyses, like ordinations, are powerful tools for presenting similarities and differences among stream sites (e.g., Wright and Li 2002) and even for examining restoration (e.g., Höckendorf et al. 2017, Manfrin et al. 2019), the wide variety of analyses used restricts comparisons among studies and the statistical complexity may hinder interpretations by a broad audience. Thus, this study took an alternative approach that individually tested differences in multiple variables between two groups of sites (e.g., restored vs. unrestored) and tested whether each variable differed between two groups of sites when considered over multiple scales. Non-parametric methods were used to mitigate issues with normality, linearity, and the need to transform or modify variables to meet those assumptions.

Differences in 42 variables were examined for each site group comparison. Eighteen variables pertained to density and size of regionally ubiquitous fish species, six variables pertained to relative abundance of the more common fish species, and 18 variables pertained to habitat characteristics. Three of the habitat variables dealing with substrate proportions were combinations of two individual variables: fines + sand, gravel + cobble, and large rock + boulder. Despite obvious correlations with the individual variables, these combination substrate variables were intended to examine differences in broader categories of substrate sizes that otherwise might be missed. With the exception of fish species variables, which were based on a single pass, all measured variables were averaged for each site from each sample period.

Differences in variable means between site groups were analyzed using non-parametric randomization tests (e.g., Fisher 1935, Edgington 1964, 1969, Edgington and Onghena 2007, Manly 2018; R version 3.6.0; R Core Team 2019). Randomization (aka permutation or resampling) tests relax requirements of large sample sizes, independence, random sampling, homoscedasticity, and balanced numbers of sample units in comparison groups (e.g., Fortin and Jacquez 2000, Legendre and Legendre 1998, Manly 2018); and all probability comes from the sampled data themselves (Edgington and Onghena 2007, Manly 2018). Randomization tests randomly shuffled the observed data for each variable across the two comparison groups and determined a difference of the means for each shuffle. In other words, for a given variable, data from two comparison groups were combined and then randomly assigned to the two groups regardless of their original group identity. After 10,000 shuffles (i.e., random permutation iterations), a frequency distribution of all the mean differences was determined. The actual observed difference of the groups’ means was compared to this distribution to determine a probability (expressed as a P value) of that difference occurring at random. Thus, a probability at the 0.05 level indicated there was a 5% chance that the actual difference between groups would have occurred at random.

Five collections of site group comparisons were considered: (1) Blue River watershed sites vs. Grant-Platte watershed sites; (2) all restored sites vs. all unrestored sites, (3) restored sites vs. pre-treatment surveys, (4) restored sites vs. control sites, and (5) pre-treatment surveys vs. control sites (Table 1). To examine differences between watersheds (collection 1), all surveys of unrestored sites (pre-treatment surveys + controls) were compared between watersheds, as were all surveys of restored sites. For collections 2–5, different nested hierarchical scales were examined with all relevant site surveys from (a) both watersheds combined; (b) within the Blue River watershed; (c) Blue River mainstem and proximate tributaries; (d) only contiguous Blue River mainstem (BR2-BR12); (e) Blue River sites BR3, BR4, BR5; (f) Blue River sites BR3, BR4, BR5 only with survey periods in common; (g) within the Grant-Platte watershed; and, (h) only Borah Creek and McPherson Branch (Table 1; Appendix S1: Fig. S1).

A scale (a–h) within any comparison collection (2–5) was examined only if there were enough data to allow for a minimum of 2000 randomization iterations. All reported comparisons enabled 10,000 iterations except for collection 2, scale f, that allowed for 2880 iterations. All scales a–h were applicable for collection 2 (Table 1). Scales
a–e met the criteria for comparison collections 4 and 5, and a–d met the criteria for comparison collection 3 (Table 1).

A variable was identified as different between site groups if the probability from the randomization test of means was less than 0.05. Although this is an ecologically arbitrary threshold that ignores potential differences at higher probabilities (Steel et al. 2013, Amrhein et al. 2019), for this project the specific threshold level is less important than the consistent application of a given threshold across all comparisons. Consistent application of a threshold enabled quantification of the number of variable differences within, and across, scales. This approach also allowed for identification of individual variables that more consistently differed amongst comparisons.

RESULTS

Watershed comparisons

More variable means differed comparing restored site surveys between the Blue River and Grant-Platte watersheds than comparing unrestored site surveys between the two watersheds (Fig. 4; Appendix S2: Table S1). Nine variable mean differences were common between the unrestored and restored comparisons, six of which were habitat related (Appendix S2: Table S1). Surveys of both restored and unrestored sites in the Blue River averaged lower adult brown trout density, greater sculpin density and proportions, larger channel dimensions, more sand-like substrates, and less combined gravel and cobble from those in the Grant-Platte watershed (randomization tests $P < 0.05$; Appendix S2: Table S1). Unrestored sites were further distinguished between watersheds regarding non-game fish abundance, embeddedness, and gravel substrates (Appendix S2: Table S1). The unique variable mean differences comparing restored site surveys included: YOY trout density and proportions, total trout density, adult trout length, rainbow trout density, adult brown trout proportions, and larger-sized substrates (Appendix S2: Table S1).

Restored vs. unrestored comparisons

All restored vs. all unrestored sites.—When data from both watersheds were combined, 18 variable means differed between all restored and all
unrestored (pre-treatment surveys + control) sites (Fig. 5; Appendix S2: Table S2). Restored sites had larger average trout sizes, more sculpins, but fewer other non-trout species of fish than unrestored sites (randomization tests \( P < 0.05; \) Appendix S2: Table S2). Restored sites also had greater average depth, lower width:depth ratios, higher proportions of large substrates, and less overhead vegetation cover and bank erosion (randomization tests \( P < 0.05; \) Appendix S2: Table S2).

Scaling down first to the Blue River watershed, again 18 variable means differed between restored and unrestored sites (Fig. 5; Appendix S2: Table S2). However, five of those differences were not observed from the combined watershed comparison. The most differences among variables (24) were observed comparing surveys of all restored and all unrestored sites from the Blue River mainstem and its proximate tributaries (Fig. 5; Appendix S2: Table S2). At this scale differences appeared throughout both fish community and habitat variables. As the scale was reduced to contiguous, mainstem Blue River sites, variables that differed decreased to 19; only three of which were not also observed at the previous scale (Fig. 5; Appendix S2: Table S2). Only nine differences were observed comparing the three contiguous sites BR3, BR4, and BR5; six of which were habitat variables (Fig. 5; Appendix S2: Table S2). When data from the three sites were further constrained to three common sampling periods, just two variables differed between restored and unrestored sites (Fig. 5; Appendix S2: Table S2).

Across all five scales of consideration within the Blue River watershed, width:depth ratio and overhead vegetation cover were reduced in restored sites (randomization tests \( P < 0.05; \) Appendix S2: Table S2). Seven common differences occurred among the four largest scales; restored sites had higher densities and proportions of adult brown trout, greater average depths, reduced width:depth ratios, more large substrates, and less overhead cover and bank erosion (randomization tests \( P < 0.05; \) Appendix S2: Table S2). At the three largest scales in the Blue River watershed, 14 common differences occurred between restored and unrestored sites (Appendix S2: Table S2).
In contrast to Blue River, the distribution and number of sites and surveys in the Grant-Platte watershed limited scale comparisons. At the watershed scale, just seven variable means differed between all Grant-Platte watershed restored and unrestored sites (Fig. 6; Appendix S2: Table S3). Three of those differences were also observed from the combined watershed comparison. Scaling down to only sites from Borah Creek and McPherson Branch, just two substrate variables differed between restored and unrestored sites (Fig. 6; Appendix S2: Table S3).

Restored sites vs. pre-treatment surveys.—Comparisons between restored sites and only pre-treatment surveys were notably similar across the four applicable scales (Fig. 7; Appendix S2: Table S4). The number of differences observed between restored sites and pre-treatment surveys sites increased with decreasing scales from 15 when watersheds were combined to 21 variables considering only contiguous, mainstem Blue River sites (Fig. 7; Appendix S2: Table S4). There were 14 variables that differed across all four scales, eight of which were habitat variables including macrophyte cover (Appendix S2: Table S4). Increased density and proportions of YOY trout within restored sites was unique (throughout the study) to contiguous mainstem Blue River site comparisons (randomization tests $P < 0.05$; Appendix S2: Table S4).

Restored sites vs. control sites.—The number of restored sites differed from the previous comparison set because some restored sites did not have pre-treatment data. Comparisons between all restored sites and control sites revealed more differences within five scales than any other collection of comparisons (Fig. 8; Appendix S2: Table S5). The most differences amongst variables (31) were observed from the Blue River mainstem and proximate tributaries (Fig. 8). Despite comparatively high numbers of differences, only four variables (depth, width:depth ratio, overhead cover, and bank erosion) were consistently different across all five scales.

Fig. 5. Summary of non-parametric randomization tests comparing 42 variable means between all restored sites and all unrestored sites (pre-treatment surveys + controls) at various scales, with focus on Blue River watershed. The number of differences counts those variables where the probability of the mean difference being random was <0.05. (See Table 1 for $n$ values of all comparisons.).
Fig. 6. Summary of non-parametric randomization tests comparing 42 variable means between all restored and all unrestored sites (pre-treatment surveys + controls) at various scales, with focus on Grant-Platte watershed. The number of differences counts those variables where the probability of the mean difference being random was <0.05. (See Table 1 for n values of all comparisons.).

Fig. 7. Summary of non-parametric randomization tests comparing 42 variable means between restored and pre-treatment surveys at various scales. The number of differences counts those variables where the probability of the mean difference being random was less than 0.05. (See Table 1 for n values of all comparisons.).
Ten variables differed across four scales, although not consistently across the same scales like with the previous comparison collection (Appendix S2: Table S5). Trends of variables.—Trends of differences (greater than or less than) between restored and unrestored sites were remarkably consistent across the three comparison collections and across scales within collections (Appendix S2: Tables S2–S5). Width:depth ratio was the most robust variable differing in all but one of the possible 17 comparisons between restored and unrestored sites (Appendix S2: Tables S2–S5). Four additional habitat variables (depth, overhead cover, bank erosion, and boulder + large rock) differed at least 13 times amongst all restored and unrestored comparisons (Appendix S2: Tables S2–S5). Fish-related variables differed less frequently across the three comparison collections and across scales within collections (Appendix S2: Tables S2–S5). Density of brown trout >100 mm each differed 11 times (Appendix S2: Tables S2–S5). Volume was the only variable with alternate trends within a given comparison collection (Appendix S2: Tables S2–S5); proportions of gravel and gravel + cobble were the only variables with alternate trends across comparison collections (Appendix S2: Tables S2–S5).

Unrestored comparisons
There were relatively few differences (ranging from 4 to 11) and limited consistency across five scales comparing pre-treatment surveys vs. control sites (Fig. 9; Appendix S2: Table S6). Most differences were observed with habitat variables (over half of them substrate related), and three scales had no differences regarding fish densities and size (Fig. 9; Appendix S2: Table S6). In pre-treatment surveys, overhead vegetation cover was less at all scales; whereas, combined cobble + gravel substrates were proportionally greater in the four Blue River watershed scales (randomization tests $P < 0.05$; Appendix S2: Table S6).

Fig. 8. Summary of non-parametric randomization tests comparing 42 variable means between restored and control sites at various scales. The number of differences counts those variables where the probability of the mean difference being random was $<0.05$. (See Table 1 for $n$ values of all comparisons.)
DISCUSSION

A common theme in stream restoration assessment literature is the overall variability and uncertainty surrounding the outcomes of stream restoration efforts. One reason for this could be the high degree of variability in the scales at which restorations are conducted and analyzed (Wohl et al. 2015, Friberg et al. 2016). Examining nested subsets, representing different scales, from the same set of ecological data may illustrate how changing spatial and/or temporal contexts influence interpretations of those data (e.g., Wright and Li 2002, Ruegg et al. 2016). Variations of this approach assessing stream restoration have identified that time since restoration (Hasselquist et al. 2015) or longitudinal position (Manfrin et al. 2019) can alter interpretations of differences between restored and unrestored sites. This study also provides a nested-scale perspective on stream restoration assessment but deviates from traditional multivariate approaches. Independently comparing whether individual variable means differed at various scales underscored the importance of both context and variable choice when assessing stream restoration.

The hypothesis that fewer variable means would differ between restored and unrestored sites at larger scales was not fully supported. Most differences occurred at large to intermediate scales in the Blue River watershed, with relatively few differences between restored and unrestored sites at the smallest scales of consideration. One possible reason for this trend is a potential cumulative impact of restoration, throughout the contiguous properties along the Blue River, that would most likely be expressed at the large to intermediate scales of consideration. Another reason for this trend could simply be loss of statistical validity. However, validity of randomization tests is less dependent on large sample sizes, or balanced numbers of sample units in comparison groups because all probability comes from random shuffles of the sampled data themselves (sensu Edgington and Onghena 2015).
2007, Manly 2018). Nevertheless, low numbers of survey periods and sites did restrict detection of differences between restored and unrestored sites (e.g., only common survey periods for BR3, BR4, BR5 in the Blue River watershed; McPherson and Borah sites in the Grant-Platte watershed).

Restoration assessment data limited in space and time are likely overshadowed by episodic intra- and inter-annual environmental variability (e.g., storm events), and/or similarities in habitat and fish communities resulting from proximity. In contrast, involving greater numbers of survey periods and sites at larger scales of consideration within a given watershed may account for both episodic environmental variability and proximity similarities, allowing differences associated with restoration to be more readily observed. However, as comparisons between restored and unrestored sites expanded beyond the Blue River watershed fewer variable means differed, suggesting inherent variability among stream reaches may indeed obfuscate differences associated with restoration at the largest scales (sensu Wiens 1989, Roni et al. 2005, Rüegg et al. 2016). This variability at the largest scales was further illustrated by comparing sites between the two adjacent watersheds. Particularly noteworthy was the degree to which the seven surveys of restored sites in the Grant-Platte watershed differed from the 57 surveys of restored sites from the Blue River watershed. Given watershed proximity and similarities in geology, land-use, and restoration approaches, it was unexpected that restored sites from the two watersheds would have more differences than the unrestored sites. Similar restoration approaches would presumably result in sites more alike than different. In addition, differences between restored sites from the two watersheds did not appear to trend consistently toward one watershed or the other. Such results highlight underlying variability that can exist when assessing the effects of restoration amongst sites from multiple watersheds.

Acknowledging that the scales of consideration will impact interpretations of stream restoration assessments should not be a deterrent to assessing restorations. Rather, it should encourage assessments to examine results across multiple, nested scales when possible. That said, care should be taken when comparing conclusions from assessments that were conducted at different scales; meta-analyses of data from different regions or watersheds provide unique (not better or worse) perspectives from those assessments done at the reach scale within the same stream. Efforts should also continue to identify robust variables that are measurable, applicable, and comparable across systems.

Habitat variable differences between restored and unrestored sites were more consistent across scales than fish community variables. Additionally, only habitat variables differed at the smallest scales of consideration. Such observations reinforce the suggestion that population data may be problematic in assessments (e.g., Friberg et al. 2016, Rubin et al. 2017), especially at smaller scales. Given differences in habitat and communities among streams, using quantitative variables of individual species and specific habitat characteristics to compare restoration projects from different regions is not applicable (Rubin et al. 2017). However, when scales and statistical analyses are similar between projects, comparing the trends of targeted variables would be warranted. For example, if two goals of restoration were to increase game species of fish and provide narrower, deeper stream channels, then the monitored trends between restored and unrestored sites could be compared regardless of the actual species or channel size. In Blue River restored sites vs. unrestored sites, consistent trends of higher average densities and proportions of the dominant game species (S. trutta) and reduced width:depth ratios would suggest restoration was achieving the two goals, and more importantly, those trends could be compared with other such projects.

Caution should be taken using trends in biodiversity to assess restoration. While a global perspective to preserve and increase biodiversity is essential, a goal of increasing biodiversity may be inappropriate for all streams and/or communities (Rubin et al. 2017). In North America, coldwater trout streams often have fewer species of fish than cool or warm water streams (e.g., Lyons et al. 2009, Myers et al. 2017). Degradation of habitat may lead to a shift toward a more diverse community that takes advantage of the new conditions. In such a case, the goal of restoration may be reducing fish community diversity. In Southwest Wisconsin, non-salmonid fish species...
were typically fewer in restored sites when differences with unrestored sites were apparent. This trend toward decreasing biodiversity could be interpreted as a positive outcome of restoration.

Ultimately, the outcomes of any restoration assessment will be dependent on the sites chosen for comparisons and the data considered. The selection of unrestored sites for comparison is both critical and challenging (Roni et al. 2005); there is no one-size-fits-all strategy. When restoration occurs within a greater landscape otherwise unaltered by humans, use of pre-disturbance, reference, unrestored sites would seem preferable. However, in landscapes dominated by a legacy of human land-use with few remaining natural areas, finding an undisturbed reference stream, or even a reach, is unlikely (Mant et al. 2016). Instead, pre-treatment surveys and control sites that have remained degraded and unrestored for multiple years provide conditions of what restored sites should be different from (e.g., Moerke et al. 2004, Roni et al. 2005, Paillex et al. 2017).

This project considered both pre-treatment surveys and control sites for comparison. As expected, there were comparatively few variables that differed between pre-treatment surveys and control sites that remained unrestored throughout the study period. Given limited differences between the two types of unrestored sites (especially with respect to fish communities), it might be assumed that that each type of unrestored sites should perform similarly in comparisons with restored sites. However, when separately compared with restored sites there were marked differences.

Fewer overall differences comparing restored sites to their pre-treatment surveys is likely indicative of the inherent similarities that exist simply being the same sites. When this inherent similarity is removed by comparing restored sites to only control sites, more differences were observed. Also, the variables that differed were more consistent across scales when comparing restored sites to their pre-treatment surveys. While this is likely a function of examining mostly the same set and number of sites from one scale to the next, the lack of consistency across scales when comparing restored sites to only control sites emphasizes the impact that choice of comparison sites may have on interpretations. Comparing the two different collections of unrestored sites may also have illuminated a conscious, practical bias in the selection of sites that underwent restoration. Consistently lower amount of overhead vegetation cover in pre-treatment surveys was likely indicative of the pre-existing angling usage (i.e., less vegetation, less casting snags) and consequent restoration potential of those areas vs. non-targeted, degraded, control sites.

Data from this project were admittedly not ideal. All sites were not sampled at all times. Some sites have only a single sample period. Pre-treatment conditions were rarely monitored for more than two sample periods at any given restored site. And the two different watersheds were not equally represented. That said, the data do represent a variety of 44 sites over a 14-year span, many sites were sampled multiple times, some sites were sampled more than once in the same year, both contiguous and independent sites were surveyed, and a breadth of biotic and abiotic variables was considered. While not ideal, the dataset does mitigate some of the common pitfalls identified in other assessments by Rubin et al. (2017).

This project should be considered an initial, alternative attempt to address lingering issues regarding restoration assessments, not a definitive conclusion. Instead of using more traditional statistical analyses to provide a consolidated result (e.g., an ordination plot with vectors, MANOVA results table), non-parametric randomization tests were used to provide an array of results for a new perspective. Though used in many other areas of biology (Manly 2018), randomization tests are missing, or rare, among stream restoration assessments. Although mean differences were used for comparisons, the same analysis could be done with medians, standard error, or other variable metrics. Alternative examinations of effect size and/or factor analysis also would provide insight beyond the scope of this initial study. While the same variables were used consistently across all comparisons and each variable was tested independently of the other variables, with stream comparison studies, independence, replication, and variable correlation should be acknowledged regarding any conclusions.
Roni et al. (2005) argued stream restoration assessments should be reported regardless of success or failure. However, determining success or failure depends on the project goals and assessment criteria which may vary greatly among restoration projects (e.g., Wohl et al. 2005, Thomas et al. 2015). If the ultimate goal of restoration is “assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed” (Society for Ecological Restoration, What is ecological restoration?, https://www.ser-rrc.org/what-is-ecological-restoration), it would seem the most basic assessment criterion for all restorations is simply that there are observable, protracted, goal-directional differences between restored areas and unrestored areas. Within that framework, finding assessment approaches that are transferable and comparable would help address challenges associated with scale, context, and variable selection. This work from Southwest Wisconsin is but one more step in that direction.

ACKNOWLEDGMENTS

Thanks to my amazing team of undergraduates who helped collect data over the years: Amanda Zopp, Brad Krykendall, Elliott Reams, Katherine Lidbury, Jenna Peterson, Jennie Franks, Robert Burken III, Russ Wolf, Sarah Scott, Patrick Gellings, Emily Hastings, Kelsey Hagan, Eric Goebel, Kailee Fiorello, Logan Steinhoff, and Mackenzie Schuenemann. Thanks to the Harry and Laura Nohr Chapter of Trout Unlimited for funding and collaborative support. Austin Polebitski provided a great deal of programming assistance for the data analysis with R. Comments from Ryan Haasl, Evan Larson, Becky Doyle-Morin, and various colleagues improved the manuscript. Thanks to the editorial staff and anonymous reviewers for their assistance.

LITERATURE CITED

Amrhein, V., S. Greenland, and B. McShane. 2019. Scientists rise up against statistical significance. Nature 567:305–307.

Benedetti, M. M. 1993. Sediment budget response to land use changes: Big Jack Watershed, Grant County. Thesis. University of Wisconsin-Madison, Madison, Wisconsin, USA.

Edgington, E. S. 1964. Randomization tests. Journal of Psychology 57:445–449.

Edgington, E. S. 1969. Approximate randomization tests. Journal of Psychology 72:143–149.

Edgington, E. S., and P. Onghena. 2007. Randomization tests. Fourth. Chapman and Hall/CRC, New York City, New York, USA.

Fisher, R. A. 1935. The design of experiments. Oliver and Boyd, Edinburgh, UK.

Fortin, M. J., and G. M. Jacquez. 2000. Randomization tests and spatially auto-correlated data. Bulletin of the Ecological Society of America 81:201–205.

Friberg, N., N. V. Angelopoulos, A. D. Buijse, I. G. Cowx, J. Kail, T. F. Moe, H. Moir, M. T. O’Hare, P. F. M. Verdonschot, and C. Wolter. 2016. Effective river restoration in the 21st century: from trial and error to novel evidence-based approaches. Advances in Ecological Research 55:535–611.

Hasselquist, M., C. Nilsson, J. Häxltén, D. Jorgensen, L. Lind, and L. E. Polvi. 2015. Time for recovery of riparian plants in restored northern Swedish streams: a chronosequence study. Ecological Applications 25:1373–1389.

Höckendorff, S., J. D. Tonkin, P. Haase, M. Bunzel-Druke, O. Zimball, M. Scharf, and S. Stoll. 2017. Characterizing fish responses to a river restoration over 21 years based on species’ traits. Conservation Biology 31:1098–1108.

Hunt, R. L. 1993. Trout stream therapy. University of Wisconsin Press, Madison, Wisconsin, USA.

Jähning, S. C., A. W. Lorenz, D. Hering, C. Antons, A. Sundermann, E. Jedicke, and P. Haase. 2011. River restoration success: a question of perception. Ecological Applications 21:2007–2015.

Juckem, P. F., R. J. Hunt, M. P. Anderson, and D. M. Robertson. 2008. Effects of climate and land management change on streamflow in the Driftless Area of Wisconsin. Journal of Hydrology 355:123–130.

Kail, J., K. Brabec, M. Poppe, and K. Januschke. 2015. The effect of river restoration on fish, macroinvertebrates and aquatic macrophytes: a meta-analysis. Ecological Indicators 58:311–321.

Knox, J. C. 2001. Agricultural influence on landscape sensitivity in the Upper Mississippi River Valley. Catena 42:193–224.

Legendre, P., and L. Legendre. 2012. Numerical ecology. Third edition. Elsevier, Oxford, UK.

Lyons, J., B. M. Weigel, L. K. Paine, and D. J. Under- sander. 2000. Influence of intensive rotational grazing on bank erosion, fish habitat quality, and fish communities in southwestern Wisconsin trout streams. Journal of Soil and Water Conservation 55:271–276.

Lyons, J., T. Zorn, J. Stewart, P. Seelbach, K. Wehrly, and L. Wang. 2009. Defining and characterizing coolwater streams and their fish assemblages in Michigan and Wisconsin, USA. North American Journal of Fisheries Management 29:1130–1151.
Manfrin, A., S. Teurlincx, A. W. Lorenz, P. Haase, M. Marttila, J. T. Syrjänén, G. Thomas, and S. Stoll. 2019. Effect of river restoration on life-history strategies in fish communities. Science of the Total Environment 663:486–495.

Manly, B. F. J. 2018. Randomization, bootstrap and Monte Carlo methods in biology. Third edition. Chapman and Hall/CRC, New York City, New York, USA.

Mant, J., A. Large, and M. Newson. 2016. River restoration: from site-specific rehabilitation design towards ecosystem-based approaches. Pages 313–334 in D. J. Gilvear, M. T. Greenwood, M. C. Thoms, and P. J. Wood, editors. River science: research and management for the 21st century. John Wiley and Sons, New York City, New York, USA.

Mauldin, L., and J. Hastings. 2014. Driftless Area restoration effort: aquatic conservation plan. Trout Unlimited, Madison, Wisconsin, USA.

Moerke, A. H., K. J. Gerard, J. A. Latimore, R. A. Helmthal, and G. A. Lambert. 2004. Restoration of an Indiana, USA, stream: bridging the gap between basic and applied lotic ecology. Journal of the North American Benthological Society 23:647–660.

Myers, B. J. E., A. J. Lynch, D. B. Bunnell, C. Chu, J. A. Falke, R. P. Kovach, T. J. Krabbenhoft, T. J. Kwak, and C. P. Paukert. 2017. Global synthesis of the documented and projected effects of climate change on inland fishes. Reviews in Fish Biology and Fisheries 27:339–361.

Paillex, A., N. Schuwirth, A. W. Lorenz, K. Januschke, A. Peter, and P. Reichert. 2017. Integrating and extending ecological river assessment: concept and test with two restoration projects. Ecological Indicators 72:131–141.

Palmer, M. A., et al. 2005. Standards for ecologically successful river restoration. Journal of Applied Ecology 42:208–217.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Roni, P. 2019. Does river restoration increase fish abundance and survival or concentrate fish? The effects of project scale, location, and fish life history. Fisheries 44:7–19.

Roni, P., M. C. Liermann, C. Jordan, and E. A. Steel. 2005. Steps for designing a monitoring and evaluation program for aquatic restoration. Pages 13–34 in P. Roni and E. Quimby, editors. Monitoring stream and watershed restoration. American Fisheries Society, Bethesda, Maryland, USA.

Rubin, Z., G. M. Kondolf, and B. Rios-Touma. 2017. Evaluating stream restoration projects: what do we learn from monitoring? Water 9:174.

Rüegg, J., et al. 2016. Baseflow physical characteristics differ at multiple spatial scales in stream networks across diverse biomes. Landscape Ecology 31:119–136.

Rydgren, K., R. Halvorsen, J. P. Töpper, I. Auestad, L. N. Hamre, E. Jongejans, and J. Sulavik. 2019. Advancing restoration ecology: a new approach to predict time to recovery. Journal of Applied Ecology 56:225–234.

Shirey, P., M. Brueseke, J. Kenny, and G. Lamberti. 2016. Long-term fish community response to a reach-scale stream restoration. Ecology and Society 21:11.

Simaika, J. P., S. Stoll, A. W. Lorenz, G. Thomas, A. Sundermann, and P. Haase. 2015. Bundles of stream restoration measures and their effects on fish communities. Limnologia 55:1–8.

Steel, E. A., M. C. Kennedy, P. G. Cunningham, and J. S. Stanovick. 2013. Applied statistics in ecology: common pitfalls and simple solutions. Ecosphere 4:1–13.

Stoll, S., J. Kail, A. W. Lorenz, A. Sundermann, and P. Haase. 2014. The importance of the regional species pool, ecological species traits and local habitat conditions for the colonization of restored river reaches by fish. PLOS ONE 9:e84741.

Stoll, S., A. Sundermann, A. W. Lorenz, J. Kail, and P. Haase. 2013. Small and impoverished regional species pools constrain colonisation of restored river reaches by fishes. Freshwater Biology 58:664–674.

Sundermann, A., S. Stoll, and P. Haase. 2011. River restoration success depends on the species pool of the immediate surroundings. Ecological Applications 21:1962–1971.

Thomas, G., A. W. Lorenz, A. Sundermann, P. Haase, A. Peter, and S. Stoll. 2015. Fish community responses and the temporal dynamics of recovery following river habitat restorations in Europe. Freshwater Science 34:975–990.

Thorn, W. C., C. S. Anderson, W. E. Lorenzen, D. L. Hendrickson, and J. W. Wagner. 1997. A review of trout management in southeast Minnesota streams. North American Journal of Fisheries Management 17:860–872.

Tonkin, J. D., S. Stoll, A. Sundermann, and P. Haase. 2014. Dispersal distance and the pool of taxa, but not barriers, determine the colonisation of restored river reaches by benthic invertebrates. Freshwater Biology 59:1843–1855.

Trimble, S. W. 1993. The distributed sediment budget model and watershed management in the Paleozoic Plateau of the upper Midwestern United States. Physical Geography 14:285–303.
Vetrano, D. M. 1988. Unit construction of trout habitat improvement structures for Wisconsin coulee streams. Administrative Report No. 27. Bureau of Fisheries Management, Wisconsin Department of Natural Resources, Madison, Wisconsin, USA.

Weber, C., U. Åberg, A. D. Buijse, F. M. R. Hughes, B. G. McKie, H. Piégay, P. Roni, S. Vollenweider, and S. Haertel-Borer. 2018. Goals and principles for programmatic river restoration monitoring and evaluation: collaborative learning across multiple projects. Wiley Interdisciplinary Reviews: Water 5: e1257.

Wentworth, C. K. 1922. A scale of grade and class terms for clastic sediments. Journal of Geology 30:377–392.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Wisconsin Department of Natural Resources. 2002. Guidelines for evaluating habitat of wadable streams. General Technical Report NC-164. USDA Forest Service, Madison, Wisconsin, USA.

Wohl, E., S. N. Lane, and A. C. Wilcox. 2015. The science and practice of river restoration. Water Resources Research 51:5974–5997.

Wohl, E., P. L. Angermeier, B. Bledsoe, G. M. Kondolf, L. MacDonnell, D. M. Merritt, M. A. Palmer, N. L. Poff, and D. Tarboton. 2005. River restoration. Water Resources Research 41:W10301. https://doi.org/10.1029/2005WR003985

Wright, K. K., and J. L. Li. 2002. From continua to patches: examining stream community structure over large environmental gradients. Canadian Journal of Fisheries and Aquatic Sciences 59:1404–1417.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3440/full