Meta-analysis of the effects of upstream land cover on stream recovery

Bronwen Stanford 1,*, Holly Jones 2, and Erika Zavaleta 3

1Environmental Studies Department, University of California Santa Cruz, 1156 High St., Santa Cruz, CA 95064, U.S.A.
2Department of Biological Sciences and Institute for the Study of the Environment, Sustainability, & Energy, Northern Illinois University, Montgomery 448, DeKalb, IL 60115, U.S.A.
3Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 1156 High St., Santa Cruz, CA 95064, U.S.A.

Abstract: Unpredictable or variable ecosystem recovery from disturbance presents a challenge to conservation, particularly as the scale of human disturbance continues to increase. Theory suggests land-cover and disturbance characteristics affect recovery, but individual studies of disturbance and recovery frequently struggle to uncover generalizable patterns because of high levels of site-specific variation. To understand how land-cover, disturbance type, and disturbance duration influence ecosystem recovery, we used studies documenting recovery of 50 streams to perform a global meta-analysis of stream recovery from disturbances that affect water quality (e.g., oil spill, fire, wastewater). We extracted upstream natural and urban land-cover percentages for each site and performed model selection and averaging to identify influences on recovery completeness. Most streams improved following the end of a disturbance (median 240% of disturbed condition) but did not recover fully to baseline predisturbance condition within the studied period (median study period 2 years; median recovery 60% of baseline). Scale of disturbance in time and space did not predict recovery, but sites with higher percentages of upstream natural land cover had less complete recovery relative to sites with more urban or agricultural cover, possibly due to higher baseline conditions in these streams. Our findings suggest impacts to systems with low anthropogenic stress may be more irreversible than impacts to already modified systems. We call for more long-term evaluations of ecosystem response to disturbance and the inclusion of regional references and predisturbance reference conditions for comparison. A more thorough understanding of the role of the surrounding landscape in shaping stream response to disturbance can help managers calibrate expectations for recovery and prioritize protection.

Keywords: anthropogenic change, baseline, bioassessment, biotic assemblages, disturbance scale, ecosystem recovery

Meta-Análisis de los Efectos de la Cobertura Río-Arriba sobre la Restauración de Arroyos

Resumen: La restauración impredecible o variable de un ecosistema después de una perturbación presenta un reto para la conservación, particularmente conforme la escala de perturbaciones humanas continúa incrementándose. La teoría sugiere que la cobertura de suelo y las características de la perturbación afectan a la restauración pero los estudios individuales sobre las perturbaciones y las restauraciones constantemente luchan por descubrir patrones generalizables debido a los niveles altos de variación específica en el sitio. Usamos estudios que documentan la restauración de 50 arroyos para realizar un meta-análisis global de la restauración de arroyos después de perturbaciones que afectaron la calidad del agua (p. ej.: derrames de petróleo, incendios, aguas negras) y así entender cómo la cobertura de suelo, el tipo de perturbación, y la duración de la perturbación influyen sobre la restauración del ecosistema. Extraímos porcentajes de la cobertura natural y urbana de suelo río-arriba para cada sitio y realizamos una selección y promedio de modelos para identificar las influencias sobre la completitud de la restauración. La mayoría de los arroyos mejoraron después de que terminó la perturbación (mediana del 240% de la condición perturbada) pero no

*email: bronwenstanford@gmail.com

Article impact statement: Understanding relative site vulnerability to disturbance can calibrate expectations and assist with prioritizing sites for protection.

Paper submitted April 11, 2018; revised manuscript accepted July 10, 2018.
se recuperó completamente hasta la línea base de condiciones previas a la perturbación dentro del periodo estudiado (mediana del periodo de estudio: dos años; mediana de la restauración 60% de la línea base). La escala de perturbación en el tiempo y en el espacio no pronosticó la restauración, pero los sitios con porcentajes más altos de cobertura natural de suelo río-arriba tuvieron una restauración menos completa en relación con los sitios con una cobertura más urbana o agrícola, posiblemente debido a las condiciones más altas de línea base en estos arroyos. Nuestros hallazgos sugieren que los impactos sobre los sistemas con un bajo estrés antropogénico pueden ser más irreversibles que los impactos sobre sistemas que ya han sido modificados. Hacemos un llamado por más evaluaciones a largo plazo de la respuesta de los ecosistemas ante las perturbaciones y por la inclusión de referencias regionales y condiciones previas a la perturbación como referencia para realizar comparaciones. Un entendimiento más a fondo del papel del paisaje circundante en la formación de la respuesta de los arroyos ante las perturbaciones puede ayudar a los administradores a calibrar expectativas para la restauración y a priorizar la protección.

**Palabras Clave:** bio-evaluación, cambio antropogénico, ensambles bióticos, escala de perturbación, línea base, restauración de ecosistemas

**Resumen:** El impacto de la intervención en los ecosistemas es un campo de estudio importante en la gestión del suelo. La recuperación de los ecosistemas puede ser complicada y afectada por la presencia de distintos factores. En este estudio, se evaluaron los efectos de la perturbación y la recuperación en diferentes contextos. Los resultados mostraron que la recuperación es posible, pero depende de varios factores, como la intensidad de la perturbación y las condiciones ambientales. La intervención humana debe ser cuidadosa y basada en la ciencia para garantizar la recuperación de los ecosistemas.

**Introducción**

Los ecosistemas están en constante cambio debido a la intervención humana. La perturbación y la recuperación son procesos dinámicos que requieren una comprensión profunda. El conocimiento de los procesos que conducen a la recuperación es fundamental para la conservación de los ecosistemas. La recuperación puede ser una herramienta valiosa para la gestión del suelo, ya que permite compensar los impactos negativos en los ecosistemas. Sin embargo, la recuperación no es un proceso automático y requiere una intervención humana bien planificada. En este trabajo, se evalúan los efectos de la intervención en los ecosistemas y se analizan los procesos que conducen a la recuperación.

Human impacts to ecosystems continue to intensify, and managers often must choose where to allow impacts and where to prioritize ecosystem protection. Accurate expectations for recovery can help promote strategic choices. However, assessment of the relative risks and benefits is limited by incomplete knowledge of ecosystem recovery trajectories following a disturbance. Anthropogenic disturbances such as pollution, fire, and invasive species introductions can have lasting impacts on ecosystem condition, and accurate predictions for recovery following a disturbance remain elusive (e.g., McCrackin et al. 2016; Meli et al. 2017; Jones et al. 2018). Restoration and recovery are highly site specific, and there is a need for better understanding of which sites are likely to be particularly vulnerable and how recovery trajectories may vary.

Two potential influences on ecosystem recovery from a disturbance are the scale of the disturbance and the condition of site surroundings. The scale of a disturbance in time and space influences both the size of the initial impact and availability of dispersing organisms (Peterson 2002; Standish et al. 2014). For example, a landscape-scale disturbance could eliminate colonist sources much more effectively than a localized disturbance, and a disruption of long duration might result in more permanent alterations to species assemblages and ecosystem functions (Lake 2000).

Many researchers have established the influence of surrounding land cover on ecosystem condition (Fährig 2003; Leite et al. 2013), but the influence of land cover on ecosystem recovery has been difficult to quantify. More natural cover might speed recovery by creating more varied habitats that can serve as refugia (Sedell et al. 1990) or greater connectivity to other high-quality habitats to support dispersal (Holl & Aide 2011; Leite et al. 2013). Communities that are well connected to high-quality habitats may receive subsidies from their surroundings, promoting stability over time (Baxter et al. 2005). However, complete recovery could also take longer in sites with more natural cover because baseline conditions tend to be higher. Higher-quality habitats typically have greater species richness and support more sensitive species (Fährig 2003), whereas simplified communities in sites with substantial human influence may be more tolerant or quick to recover. In this case, the different baselines against which recovery is assessed in highly simplified versus higher-quality locations could create differences in the rate or completeness of recovery.

Streams are nested within catchments and are highly sensitive to surrounding land cover (Allan 2004), making...
them particularly good model systems to evaluate questions about the effect of the surrounding landscape on recovery progress. Streams in areas with high natural cover typically also contain more sensitive species, links to other patches of high-quality habitat, and complex habitat that contains refugia for aquatic organisms such as fish and invertebrates (Urban et al. 2006). Human land uses affect stream ecological community composition, with increased land use intensity typically resulting in more tolerant, less diverse biological assemblages (Allan 2004; Lorenz & Feld 2013; Roy et al. 2016). Thus, human land use is also likely to affect stream recovery and reassembly patterns.

To assess the influence of disturbance scale and land cover on stream recovery, we conducted a meta-analysis of the response of stream ecosystems to disturbance. We limited our assessment to disturbances affecting stream condition primarily through their effect on water quality (such as oil spills, wastewater discharges, and experimental pollution) because water-quality impacts occur at a range of spatial and temporal scales and are common across a broad range of landscape types, from highly natural to highly altered contexts. The influence of land cover on recovery can be difficult to determine from an individual study because of site-specific variation and the difficulty of conducting a single research project across a variety of land-cover types. Meta-analysis is a powerful tool that allows the combination of results across studies and settings to detect patterns that may be contradictory, weak, or otherwise not captured within an individual study. Although a few meta-analyses have evaluated stream recovery (e.g., Miller et al. 2010; Smucker & Detenbeck 2014; Sievers et al. 2017), only 1 included information on land-cover type. Miller et al. (2010) used coarse categories indicated by study authors to categorize land-cover types. They found that streams in forested sites did improve, but they found no significant differences among land-cover types. We created standardized assessments of surrounding land cover to explicitly evaluate whether scale of disturbance and type of surrounding land cover affects the ability of streams to reach the predisturbance or goal condition (complete recovery) following disturbances that affect water quality.

We asked the following questions: How much recovery is achieved by streams following disturbances that affect water quality? And how do surrounding land cover and scale of disturbance in time and space affect stream recovery? Together answers to these questions can help identify vulnerable sites and those that are likely to recover.

Methods

Study Selection

We conducted a search for stream recovery studies on Web of Science on September 26, 2016. We performed a topic search for papers in English with terms for stream (stream or river or aquatic or creek) recovery (recover* or restor*) from a disturbance affecting water quality (pollut* or “water quality”) worldwide. We reviewed papers covering a broad suite of disturbances that affect rivers primarily through changes to water quality, as described by study authors (e.g., mining spill, oil spill, logging). Because our primary interest was the ecological response, we also included a term for an ecological indicator (abundance or diversity or ecolog* or richness or similarity or composition), which excluded papers focused solely on recovery of water quality. Thus, the search included terms for stream + recovery + water quality + ecological and returned 1,643 papers.

We retained 29 papers for data extraction based on the following criteria (see Supporting Information): presents quantitative data on in-stream biological condition of a stream or river; documents recovery from a disturbance that affects water quality, including data from at least 2 periods after the disturbance ended; includes a reference condition (based on either a nearby site or predisturbance data at the impact site); is not a mesocosm (had to be a catchment affecting stream condition); and includes information adequate to place the site with GIS. We combined these papers with similar databases of published studies from Jones et al. (2018) and Meli et al. (2014), which were more general meta-analyses but included some studies of stream recovery from a water-quality disturbance. Our final database contained 37 studies that document 50 streams recovering from disturbances affecting water quality.

Data Extraction

For each recovered stream, we collected data describing 3 states: affected (immediately after the cessation of the disturbance), recovered (the last time point recorded by the authors), and reference (a site not subject to the disturbance). The reference condition was either the same site before the disturbance or a nearby site. Some studies followed a before–after–control–impact design and included 2 unaffected references. In those cases, we chose the same-site, predisturbance data as the reference. In cases where there were multiple reference sampling sites on the same stream, we chose the site closest to the source of water quality disturbance. In cases with multiple data collection efforts over time, we chose the earliest postdisturbance data point as affected and the last data point as recovered, while matching for season when possible. We extracted data from figures with DataThief version 1.7 (Tummers 2006). We extracted data for all response variables that the study authors used to measure recovery. Where provided within a study also describing ecological condition, we collected data on water quality and physical habitat so that we could compare these to biotic condition.
Effect Size Calculation

In meta-analysis the log response ratio is a measure of effect size, which represents the difference between the control and treatment case for the variable of interest (Gurevitch & Hedges 2001). Response ratios are a standard unitless approach to scale the values of response variables so that effect sizes for many different types of response variables can be compared (Hedges et al. 1999; Gurevitch & Hedges 2001). Response ratios do not require estimates of standard error, an important consideration, as most data in our study lacked error estimates. We calculated recovery completeness with a response ratio comparing the recovered and initial baseline state (\(\ln(\text{recovered/reference})\)). For comparison, we also calculated improvement over the perturbed condition (\(\ln(\text{recovered/affected})\)). Finally, we calculated the relative impact magnitude (\(\ln(\text{affected/reference})\)). As other recent restoration meta-analyses have done (Meli et al. 2017), we reversed the sign for variables whose values increased rather than declined under disturbance. Recovery assessments are presented in the text in back-transformed values for easier interpretation. For all plots and models, we used the original log ratios.

Eighteen percent of our data set (105 responses) included 0 values, which represented meaningful data. To avoid undefined response ratios, we added 0.01 to the numerator and denominator of all values (Supporting Information).

Predictor Variables

To characterize the surrounding landscape, we used information within each paper to spatially locate each study site and used ArcGIS online to calculate the catchment area and ESRI ArcMap to calculate land cover upstream of each site with existing stream and land-cover data sets. We calculated 2 cover types: natural cover, which we defined as all nonurban and nonagricultural land-cover types (i.e., forest, wetlands, grasslands, scrub, tree plantations) and urban cover (urban cover is frequently, particularly stressful to aquatic ecosystems [e.g., Walsh et al. 2005]). Although tree plantations may function differently from natural forests, most data sets did not distinguish the 2, so we refer to all nonurban and nonagricultural lands as natural cover. We extracted land-cover percentages at a variety of buffer widths. However, both percent natural cover and percent urban cover were consistent across scales so we selected 1 representative scale for each land-cover type (Supporting Information). The representative scale was chosen for its high correlation with other scales (natural cover, 600 m wide and 1 km upstream; urban cover, 120 m wide and 5 km upstream).

In addition to information on land cover, we collected information on study design and disturbance characteristics (Tables 1 & 2). To assess the importance of disturbance scale, we categorized the water quality disturbance in each study as either point source (from a single, identifiable point, such as a drainage pipe), or nonpoint source (a more diffuse landscape-scale impact, such as fire). We also included the log response ratio for impact magnitude (see above) as a covariate. Impact magnitude is the post-disturbance (affected) condition scaled to the reference condition. We included this term so we could account for differences in impairment to more meaningfully compare recovery across studies and expected high impact magnitude to result in low recovery completeness.

Prior to model construction, we assessed the correlation of our predictor variables with the R package corplot (Wei & Simko 2016). We found high correlation (Pearson’s \(r = -0.6\)) between catchment area and natural land cover—larger catchments had lower percentages of natural cover. To assess which variable was a better predictor, we compared the full model with natural cover as a predictor to the full model with catchment area as a predictor. Natural cover had a lower AICc score (\(\Delta\text{AICc} = 4.5\)), so we used natural cover only in our models, acknowledging that some of the effect may be due to differences in catchment area.

Meta-regression

We constructed 2 types of models. First, we used meta-regression to model both overall recovery completeness and overall improvement. We constructed a simple model for each, with response metric type as fixed effect and a site-level random effect to account for nonindependence of multiple observations from the same site (Zuur et al. 2009).

Second, we constructed a full model to assess which factors predicted recovery completeness, including response metric type, impact magnitude, percent natural cover, percent urban cover, point source versus nonpoint source, study duration, reference type, and disturbance duration as fixed effects and site as a random effect (Table 2). To enable comparison of the importance of different predictors, we standardized all continuous terms (\((x - \text{mean}[x]) / 2\text{sd}[x]\)) (Gelman 2008). Our full model contained 2 categorical variables: metric type and reference type. In both cases, we set the largest category as the reference level: abundance is the reference level of metric type and predisturbance is the reference level of reference type. We used the R package metafor to construct models (Viechtbauer 2010). All analyses were completed with R version 3.4.3 (R Core Team 2017).

We compared subsets of the full model with the bias-corrected Aikake information criterion (AICc) (Burnham & Anderson 2002) and the R package glmulti (Calcagno 2013). We forced the inclusion of metric type in every model but evaluated all combinations of the other variables. There was no clearly preferred model, so we selected models within 2 AICc values of the lowest AICc score and used these to calculate model-averaged...
Table 1. Stream-condition metrics included in the study of the effects of upstream land cover on stream recovery.a

<table>
<thead>
<tr>
<th>Metric category</th>
<th>Frequency</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abundance</td>
<td>351</td>
<td>abundance of collector gatherers</td>
</tr>
<tr>
<td>abundance</td>
<td>259</td>
<td>mean density of trout</td>
</tr>
<tr>
<td>density</td>
<td>49</td>
<td>bedrock substrate annual production of scrapers (g AFDMb/m²/yr)</td>
</tr>
<tr>
<td>production</td>
<td>31</td>
<td>habitat weighted biomass of insects (g AFDM/m²)</td>
</tr>
<tr>
<td>biomass</td>
<td>6</td>
<td>arctic grayling cohort-specific growth rate (%/day)</td>
</tr>
<tr>
<td>growth</td>
<td>5</td>
<td>bryophyte cover (%)</td>
</tr>
<tr>
<td>percent cover</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Diversity</td>
<td>98</td>
<td>number of Mollusca taxa</td>
</tr>
<tr>
<td>richness</td>
<td>79</td>
<td>Shannon-Wiener diversity of diatoms</td>
</tr>
<tr>
<td>diversity</td>
<td>15</td>
<td>% Ephemeroptera richness</td>
</tr>
<tr>
<td>Community composition</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Biotic integrity</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>health</td>
<td>14</td>
<td>fish stomach fullness</td>
</tr>
<tr>
<td>index</td>
<td>25</td>
<td>ecotoxicological rating (includes inverts, water quality)</td>
</tr>
<tr>
<td>survival</td>
<td>8</td>
<td>Asian clam % survival</td>
</tr>
<tr>
<td>Physical habitat</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>organic matter</td>
<td>35</td>
<td>volume of large wood within bankfull width (m³/m²)</td>
</tr>
<tr>
<td>sediment</td>
<td>7</td>
<td>unit channel sediment storage (m³/m²)</td>
</tr>
<tr>
<td>channel</td>
<td>3</td>
<td>mean bankfull width (m)</td>
</tr>
<tr>
<td>Water quality</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>pollution</td>
<td>20</td>
<td>copper concentration in water (mg/L)</td>
</tr>
<tr>
<td>nutrients</td>
<td>16</td>
<td>concentration of nitrate and nitrite (µg n/L)</td>
</tr>
</tbody>
</table>

aThe 5 first-order categories were included as categorical variables in our models of stream recovery; frequencies represent total data points within each category.
bAsh-free dry mass.

Table 2. Summary of terms included in the models of stream recovery, including predictor and response variables.

<table>
<thead>
<tr>
<th>Terma</th>
<th>Type</th>
<th>Transformationb</th>
<th>Method of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>metric typec</td>
<td>categorical, 5 levels</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>natural cover (%)</td>
<td>continuous</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>urban cover (%)</td>
<td>continuous</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>reference type</td>
<td>categorical, 3 levels</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>point source vs. nonpoint source</td>
<td>binary</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>disturbance duration (years)</td>
<td>continuous</td>
<td>log.</td>
</tr>
<tr>
<td></td>
<td>study duration (years)</td>
<td>continuous</td>
<td>log.</td>
</tr>
<tr>
<td></td>
<td>impact magnitude</td>
<td>response ratio</td>
<td>log.</td>
</tr>
<tr>
<td>Response</td>
<td>recovery completenessf</td>
<td>response ratio</td>
<td>log.</td>
</tr>
<tr>
<td></td>
<td>improvement</td>
<td>response ratio</td>
<td>log.</td>
</tr>
<tr>
<td></td>
<td>impact magnitude</td>
<td>response ratio</td>
<td>log.</td>
</tr>
</tbody>
</table>

aAll models also included site as a random effect.
bAll continuous terms were standardized: \[x \rightarrow \frac{\text{mean}(x)}{2\times\text{SD}(x)}\].
cMetric type was included in all models; other terms included only in full model.
dAffected refers to the earliest data taken following the end of the disturbance.
eReference refers to data taken at the reference site (either nearby, upstream, or predisturbance) and represents the predisturbance baseline.
fPrimary response variable used in this study.
gRecovered refers to latest data taken following the end of the disturbance.
coefficients weighted by the relative AICc scores (Burnham & Anderson 2002; Grueber et al. 2011). We used these model-averaged coefficients to evaluate the relative importance of different predictors. Model-averaged coefficients can be misleading where terms are collinear (Cade 2015), so we also present the full and the top model to aid interpretation (Supporting Information). We evaluated model residuals for fit, leverage, and skew; there were no influential outliers. Although funnel plots and other assessments of publication bias are typically recommended for meta-analysis, they are ineffective for random-effects models and for models with missing variance and multiple effect sizes per study, as we have here (Lajeunesse 2009).

Meta-analysis requires an estimate of within-study variance to weight studies (Gurevitch et al. 2018). However, only 96 of our 575 data points (17%) included estimates of error. Within each model, we used the known variance where possible, and we used the model to estimate a uniform variance for all points with unknown variance (Viechtbauer 2010). We then assigned this model-estimated variance to all points with missing variance prior to completing model selection. We performed 2 checks on our variance estimate. To assess model sensitivity to the estimated variance value, we reran the top-selected model with estimated sampling variance 1/5 and 2 times the model-determined value. Estimates were qualitatively the same, so we used the model-assigned values (Supporting Information). Second, because the true variance for those points where variance is unknown is not likely to be uniform across studies, we applied a cluster-robust variance correction to data grouped at the level of the study (Angrist & Pischke 2009; Viechtbauer 2010). We assumed errors are more likely to be correlated within a study due to similar data-collection methods. We present results with this corrector in the text; results both with and without are in Supporting Information. We were unable to apply the robust corrector during model selection. Instead, we applied this corrector to the top selected model (Supporting Information).

To explore whether impact magnitude also varied by land cover, we performed a secondary analysis to model impact magnitude with the same set of terms (metric type, study duration, point source versus non-point source, reference type, disturbance duration, natural cover, urban cover) and performed model selection and model averaging with the same methods described above.

**Results**

**Summary of Data Set**

The final data set included 37 studies, documenting the recovery of 50 streams, of which 26 were in the were in the United States and 11 in Europe, with the remainder in Argentina (6), Malaysia (3), Australia (2), New Zealand (1), and Canada (1). Only 3 study sites were in the tropics, and only 3 study sites had nonperennial flow.

The studies reported a total of 575 responses that met our criteria. Abundance of 1 or more types of organisms was the most common response metric (60% of responses), and macroinvertebrates were the most common taxonomic group studied (56% of responses, and measured in some way in 31 of the streams) (Supporting Information). Nine streams were affected by natural disturbances (fire, 2 streams; hurricane, 1; volcanic eruption, 6), whereas the remainder were affected by human-caused disturbances (logging, 4; experimental treatments of biocide, 3; nutrients, 2) or opportunistic studies of recovery from spills (8), wastewater discharges (7), logging (7), and mining (7). When we assessed natural cover by disturbance category, each category occurred either only in high natural cover sites or across a broad range of percent natural cover (Supporting Information).

Forty-two percent of sites (21 sites) were compared with predisturbance conditions; the remainder were compared with a nearby reference on another stream (32%) or an upstream reference on the same stream (26%). None of the sites were actively restored.

The median study duration (elapsed time between initial postdisturbance and final postdisturbance data collection) was 2 years, and study duration ranged from 20 days to 62 years postdisturbance (IQR = 1–3.8 years, mean [SD] = 5.3 [9.9]). Seven streams were monitored for >10 years and 7 for <1 year.

**Stream Recovery**

The median site recovered to 60% of reference condition when averaged across all measured responses (mean site = 60%; CI, 51–71). Although most sites did not recover completely, 30% of measured responses achieved over recovery—a final state above the reference condition. Most sites improved after the disturbance ended: the median site improved to 240% of disturbed condition (mean = 337%; CI, 236–480).

There were no significant differences between recovery of abundance and other metrics (Supporting Information), but biotic integrity, diversity, and water quality recovered to reference condition on average, whereas abundance and diversity did not (Fig. 1a). Abundance, diversity, and water quality improved significantly over the disturbed condition, whereas biotic integrity and physical habitat metrics did not (Fig. 1b). When we evaluated only fish and invertebrate abundance, neither recovered completely (Supporting Information).

**Influences on Recovery**

Recovery to the level of the predisturbance baseline (recovery completeness) was predicted by higher impact magnitude, less natural cover, and metric type (Fig. 2
Figure 1. Mean effect size by response metric type for (a) stream recovery completeness (dashed line, complete recovery) and (b) stream improvement over affected condition (dashed line, no improvement). Error bars are cluster-robust 95% CI. Numbers below metric not in parentheses are number of data points. Numbers in parentheses are number of sites. Estimates are considered different from 0 if CI does not overlap 0.

Figure 2. Model-averaged regression coefficients (whiskers, 95% CI) for the top predictors of recovery completeness in streams following a disturbance to water quality (ΔAICc = 2). Predictors are standardized so that coefficients can be compared directly. Estimates are different from zero if CIs do not overlap 0. Models also include metric type (see Supporting Information).

Figure 3. Estimated effect of land cover on stream recovery to predisturbance baseline (solid line, top selected model estimate; dashed lines, robust 95% CI). Other terms in the top-selected model included metric type, study duration, and impact magnitude.

Discussion

Overall, we found that streams improved but did not recover fully following disturbances affecting water quality. In addition, streams with higher percentages of upstream natural cover were less likely to recover to the predisturbance baseline condition but also less likely to be severely affected by the disturbance.
Measuring Stream Recovery

Overall, streams failed to recover to baseline. Other recent reviews and meta-analyses looking at other types of disturbances have found low recovery completeness in wetlands (Moreno-Mateos et al. 2012), forests (Meli et al. 2017), and multiple ecosystem types (Be-nayas et al. 2009; Jones et al. 2018). Meta-analyses and reviews of active stream restoration have similarly found either mixed results (Miller et al. 2010; Palmer et al. 2010; Smucker & Detenbeck 2014; Sievers et al. 2017) or lack of improvement (Stranko et al. 2012). We found improvement of most metrics and some cases of complete recovery: our focus on a discrete, reversible disturbance likely contributes to our more positive findings.

Measured recovery completeness might have increased if the study period had been longer. Ecosystems typically recover over decades to centuries or longer (Jones & Schmitz 2009), but studies in our database were overwhelmingly short term, with almost 50% (24 studies) lasting <2 years, and 18% lasting <1 year. In many cases, the final condition captured in these studies is likely not true recovery. However, recovery completeness did not improve with longer study duration. Instead, we found a (weak) negative relationship between study duration and recovery completeness. To better evaluate recovery, longer studies are needed, a call echoed by much of the restoration ecology literature (Bernhardt et al. 2007).

Our results also point to 3 important components of study design. First, the choice of metric type affected measured stream recovery: abundance did not recover fully, while diversity and biotic indices did, and fish recovered slightly more than invertebrates. Most metrics in our analysis related to ecological structure rather than function, which may reflect the fact that our search terms related more strongly to structure than function. A more in-depth evaluation of the recovery of ecological function might reveal different patterns.

Second, our findings show that the choice of how we calculate recovery influences the patterns we observe. Although we restricted our sample to studies that included both a degraded reference (affected site) and an unaffected reference (reference site), the degraded condition and unaffected reference condition are each frequently used individually as baselines in the restoration literature (Miller et al. 2010; Weber & Peter 2011; Morandi et al. 2014; Suganuma & Durigan 2015). These 2 different measures of the ecosystem recovery process—improvement (vs. degraded condition) and recovery completeness (vs. reference condition)—showed different patterns in our data set. For example, abundance showed high improvement and low recovery completeness. We urge researchers and practitioners to continue to use BACI and other designs with multiple reference types whenever possible to ensure capture of both improvement and progress towards the target reference condition.

Finally, our analysis was limited by data availability. Better reporting of study statistics (including sampling error and sample size) would have strengthened our ability to draw conclusions from this analysis (Gerstner et al. 2017). In addition, our study sites were heavily biased toward temperate and perennial streams; patterns for tropical or temporary streams may differ. Due to the limited available data, our results are only suggestive, and we hope that in future more detailed work will be possible to further explore these questions.

Influence of Land Cover and Disturbance Scale on Recovery

More natural cover predicted less complete recovery. We explored whether the effects of disturbance were also more severe in areas with natural cover, but found that more natural cover (or less urban cover) weakly predicted less severe effects. Together, these findings suggest that streams with more upstream natural cover are likely to be more stable over time, but may fail to recover fully. Disturbances in already degraded ecosystems are more likely to have a temporary effect, while disturbances to natural ecosystems may be more irreversible. These findings emphasize the importance of avoiding impacts to high-quality ecosystems.

The more complex and diverse communities typically found in streams with lower human impacts may explain the lower recovery and lower impacts in these systems. Stream condition varies predictably with land cover (Allan 2004; Norton et al. 2009) (but see Baker et al. 2006), with increasing richness and abundances of sensitive taxa as anthropogenic cover in a catchment decreases (Paul & Meyer 2001; Roy et al. 2003). Streams with a high-quality baseline condition are effectively held to a higher standard of recovery, with a larger complement of species (Stoddard et al. 2006). But, a larger complement of species also means that the loss of a couple of species results in a smaller measured impact magnitude. To allow direct comparison of recovered conditions across streams, studies need to include a regional reference. Researchers working in a variety of ecosystems have called for the use of a regional “quantitative optimum” reference representing best attainable regional condition (Stoddard et al. 2006; Morandi et al. 2014). The use of a regional reference condition would have allowed us to quantify the condition of each stream and separate the effects of initial condition and surrounding cover on recovery. Although a few of the studies in our database did include a regional reference (e.g., Arce et al. 2014), most did not, limiting the conclusions we could draw about ecosystem condition.

Less complete recovery in more natural areas could also occur because streams with more natural cover have
more sensitive, specialist, and rare species (Roy et al. 2003), so that recolonization may take more time or might never reassemble the original species composition. Less sensitive and more homogenous taxa are likely to be widespread in streams with high anthropogenic land use (Urban et al. 2006), and these communities may be more resilient to additional disturbance because all of the intolerant species have already been lost (Stoddard et al. 2006). In this study, natural cover was correlated with small catchment area, so some of the observed effect may also indicate that headwaters are less able to recover to baseline, possibly because they lack upstream sources of colonists.

To further explore these results, we call for studies of stream recovery across a broader variety of stream types. The effect of natural cover on recovery completeness may interact with hydrological variability, particularly for intermittent streams (Matono et al. 2012). Although our study included a range of upstream catchment area sizes, most were perennial, temperate streams, and patterns may differ by climate and flow vs regime.

The scale of disturbance in time and space did not predict recovery completeness. We expected long-lasting and large-scale disturbances (such as chronic nonpoint source water pollution) to result in less complete recovery. However, our results suggest that landscape condition and data collection methods have a larger influence on measured recovery than disturbance scale.

Our results echo the findings of others (e.g., Jones et al. 2018) that ecosystem recovery following disturbance is uncertain and often incomplete. Catchments with more natural cover may be less able to return to baseline conditions than catchments with extensive human land use and already simplified communities, so avoiding impacts in natural systems may be even more critical than in already modified streams. The limited number of studies available for this meta-analysis highlights the need for more rigorous studies of ecosystem improvement to support strategic conservation investments in the future.

Acknowledgments

We appreciate comments from K.D. Holl, S. Skikne, E. Olimpi, P. Jones, the Holl Lab, and 5 anonymous reviewers who helped substantially improve this manuscript. B.S. was funded by STAR Fellowship Assistance (agreement FP-91762101-0) awarded by the U.S. Environmental Protection Agency and by the Wells Fargo Coastal Sustainability Fellowship and the Northern California Association of Phi Beta Kappa. All authors designed the study, interpreted results, and edited the paper. B.S. collected the data, performed analysis, and wrote the paper.

Supporting Information

PRISMA diagram for study selection (Appendix S1), comparison of raw and transformed values for recovery completeness (Appendix S2), land-cover correlation matrix (Appendix S3), model sensitivity (Appendix S4), frequency of each response metric and taxonomic group (Appendix S5), type of impact to water quality plotted against % natural land cover (Appendix S6), recovery completeness of fish and invertebrate abundance (Appendix S7), data sources by region (Appendix S8), recovery completeness models with robust correction (Appendix S9), top selected models for recovery completeness (Appendix S10), top selected models for impact magnitude (Appendix S11), and papers included in the meta-analysis (Appendix S12) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited


