The effectiveness of spawning habitat creation or enhancement for substrate-spawning temperate fish: a systematic review

Jessica J. Taylor[1,2,3*†] , Trina Rytwinski[1,2,3†], Joseph R. Bennett[1,3], Karen E. Smokorowski[4], Nicolas W. R. Lapointe[5], Richard Janusz[6], Keith Clarke[7], Bill Tonn[8], Jessica C. Walsh[9] and Steven J. Cooke[1,2,3]

Abstract

Background: Habitat is the foundation for healthy and productive fisheries. For fish that require substrate for spawning, lack of appropriate spawning substrate is inherently limiting and a lack of access to suitable spawning habitat will lead to population collapse. To ensure management resources are being allocated wisely and conservation targets are being achieved, there is an increased need to consider the effectiveness of techniques to enhance or create habitat that has been lost. The aim of this systematic review was to assess the effectiveness of techniques currently used to create or enhance spawning habitat for substrate-spawning (including vegetation-spawning) fish in temperate regions, and to investigate the factors that influence the effectiveness of habitat creation or enhancement.

Methods: Searches for primary research studies on the effect of spawning habitat creation or enhancement for substrate-spawning fish were conducted in bibliographic databases, on websites and an online search engine, through evidence call-outs, social media, and Advisory Team contacts, and in the bibliographies of relevant reviews. All articles were screened at two stages (title and abstract, and full-text), with consistency checks being performed at each stage. Relevant articles were critically appraised and meta-data and quantitative data were extracted into a database. All included studies were described narratively and studies that met the criteria for meta-analysis were analyzed quantitatively.

Review findings: A total of 75 studies from 64 articles were included in this systematic review and underwent data extraction and critical appraisal. The majority of these studies were from North America (78.1%) and a large percentage (63.7%) targeted salmonids. We conducted a meta-analysis using data from 22 studies with 53 data sets. Available evidence suggests that the addition or alteration of rock material (e.g., gravel, cobble) was effective in increasing the abundance of substrate-spawning fish compared to controls, with a taxonomic bias towards salmonids (5/6 data sets). The addition of plant material (e.g., large woody debris) with or without physical alterations to the waterbody (e.g., excavation) was also effective in increasing substrate-spawning fish abundance on average compared to controls. Egg life stages (i.e., nests, redds, zygotes or developing embryos) were associated with larger increases in abundance with habitat creation or enhancement than age-0 life stages (i.e., alevin, fry, young-of-the-year). We found

---

*Correspondence: jessjtaylor16@gmail.com
†Jessica J. Taylor and Trina Rytwinski contributed equally to this manuscript
[1] Canadian Centre for Evidence-Based Conservation, Institute of Environmental Sciences and Interdisciplinary Sciences, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada
Full list of author information is available at the end of the article

Taylor *et al. Environ Evid*   (2019) 8:19

Page 2 of 31

no detectable effect of ecosystem type (lotic vs. lentic waterbodies) or time since habitat creation or enhancement on intervention effectiveness for fish abundance.

Conclusions: The synthesis of available evidence suggests that the addition or alteration of rock material (e.g., addition of gravel, substrate washing) was an effective means of enhancing spawning habitat, but results may only be applicable for salmonids. Furthermore, the synthesis suggests that on average, the addition of plant material with or without waterbody modifications was also effective at increasing fish abundance. Overall, we were limited in our ability to address many of the questions that stakeholders have regarding the circumstances under which spawning habitat creation or enhancement is effective for substrate-spawning fish. Before we can provide recommendations with a higher level of certainty, we need to improve research and reporting, and expand research focus to include a broader range of species and intervention types. We provide several recommendations aimed at researchers and practitioners to improve the quality of evidence being generated.

Keywords: Aquatic, Compensate, Evidence-based policy, In-stream structures, Lithophil, Offset, Reproduction, Restoration, Rock, Phytophil

Background

Habitat is the foundation for healthy and productive fisheries [1]. When critical habitats for fish are lost, degraded or altered, their ability to support life processes of fish may be compromised [2]. To ensure that fish habitats are appropriately managed, many jurisdictions require some form of offsetting (or compensation) for habitats that will be lost or degraded due to human developments. Several methods of habitat creation or enhancement have the potential to increase fish productivity, i.e. production rates of fish species of interest, biomass [3].

Access to and quality of spawning habitats are critical to the success and productivity of a fish population [2], especially for substrate-spawning fish [4, 5]. For these species, degradation or loss of appropriate spawning substrate, often caused by human activity, is inherently limiting and may lead to population collapse [6]. Creation or enhancement of spawning habitat is often used to mitigate or offset the destruction/degradation of spawning substrate, though may also be used to simply enhance habitat. However, to be suitable for a target species, spawning habitat must have specific properties matched to the species' ecological niche, not all of which may be known to practitioners. Thus, there is much interest in identifying the extent to which spawning habitat creation or enhancement interventions are effective at increasing population size or productivity of substrate-spawning fish. Identifying the extent to which spawning habitat creation or enhancement interventions are effective is of particular interest in temperate regions where substrate-spawning species such as salmonids (e.g., Atlantic salmon, brook trout, lake trout), centrarchids (e.g., black bass), percids (e.g., walleye), and ictalurids (e.g., brown bullhead) are common, and in jurisdictions, such as Canada, that have well-developed regulatory frameworks for habitat protection and restoration.

Some of the most common restoration or enhancement measures for spawning habitat include additions of instream structures, such as large woody debris, boulders/cobbles, logs jams, and brush bundles [7]. For example, enhanced gravel beds provide suitable spawning habitat for salmonids [8, 9] and have resulted in higher survival rates to the swim-up stage (at which point alevin swim to the surface for the first time to fill their swim bladder with air) [10]. These structures not only provide favourable habitat for juveniles [3, 11], but can also recruit and store gravel [12] which is of particular benefit to many substrate spawners.

To effectively create or enhance spawning habitat, it is important to identify habitat and environmental characteristics that influence productivity for specific species. The spawning habitats used by fishes are quite varied and factors such as temperature, depth, wave exposure, water quality, water velocity, vegetation composition, and adjacency to nursery habitat, must all be considered when attempting to design a successful habitat restoration or offset project [13, 14]. Species often have specific substrate requirements and preferences [4], and spawning behavior including nest building and guarding [4, 15]. For instance, Curry and Noakes [16] examined selection of spawning sites in brook trout (*Salvelinus fontinalis*) and demonstrated important relationships between groundwater and spawning success that varied substantially among populations across geologic regions. For example, in the Canadian Shield waters, spawning in brook trout was associated with areas of distinct discharging groundwater, whereas in southwestern Ontario, discharging groundwater was observed throughout spawning areas and nonspawning areas. Other species, such as Chinook salmon (*Oncorhynchus tshawytscha*), spawn in sites with downwelling or upwelling depending on the population. These behaviours prompted further investigation into the importance of water quality characteristics such as

dissolved oxygen and temperature [13]. However, even for species for which information on the necessary physical and chemical spawning habitat attributes exists [17], it remains difficult to re-create these attributes in the wild [18].

With accelerating habitat degradation and loss of biodiversity in aquatic systems resulting from human activity [19, 20], it is becoming ever more important to consider the effectiveness of methods to enhance degraded habitat or create new habitat. Meta-analyses and systematic reviews are valuable tools to evaluate the effectiveness of conservation interventions to inform environmental policy decisions [21]. Systematic review guidelines provided by the Collaboration for Environmental Evidence ensure that evidence syntheses are rigorous, transparent, and repeatable [22]. Here, this systematic review assesses the effectiveness of techniques currently used to create or enhance spawning habitat for substrate-spawning fish.

Topic identification and stakeholder input

In 2012, Canada's *Fisheries Act* was amended to put responsibility on proponents (e.g., persons involved with commercial developments, mineral extraction, members of the public not engaged in commercial activity, or government municipalities or ministries) to avoid and mitigate any *serious harm to fish* that are part of a commercial, recreational or Aboriginal fishery, or to fish that support such a fishery resulting from projects affecting aquatic habitat. Fisheries and Oceans Canada (DFO) updated the way they managed threats to fisheries from development projects such that if projects could not avoid or mitigate serious harm, proponents were required to develop a plan to counterbalance the residual harm using offsetting measures [23–25]. Offsetting measures will differ on a case-by-case basis; however, all must support fisheries management, balance project impacts, and generate long-term, self-sustaining benefits for the fishery [23]. Resources could be more efficiently used by critically reviewing the effectiveness of past spawning habitat creation or enhancement projects.

During the formulation of the question for this review, an Advisory Team made up of stakeholders and experts was established and consulted. For the purpose of this review, we define stakeholders as "any person or organization who can affect or may be affected by the planning, conduct, results and communication of a systematic review" (see Haddaway et al. [26] for full framework). This team included academics, staff from the Canadian Wildlife Federation (CWF), and staff from DFO, specifically the Fisheries Protection Program (FPP) and Science Branch. The Advisory Team guided the focus of this review to ensure that primary and secondary questions were both answerable and relevant, and suggested search terms to capture the relevant literature. Our systematic review is complementary to a systematic review [27] that synthesized evidence on the impact of anthropogenic structural modifications to habitats in shallow water nurseries and/or spawning grounds on fish recruitment, but is broader in scope. Though methods of habitat creation or enhancement have been studied, to our knowledge no comprehensive synthesis of evidence has been undertaken to compare the effectiveness of all relevant habitat creation or enhancements for substrate-spawning fish. Some reviews have focused on broader topics such as the effect of a physical structure and cover on fish and fish habitat [17], others focused on a specific family (e.g., salmonids; [28]), a particular habitat (e.g., streams; [29]) or only review a small number of restoration studies on a specific topic (artificial reefs in the Great Lakes [30] or for production of marine fishes [31]; instream structures for salmonids [32]). Discussions with our Advisory Team confirmed the value of systematically reviewing available literature to examine how and when habitat creation or enhancement can benefit populations of substrate-spawning fish. During the course of this review, the Advisory Team was consulted to develop the data extraction table and critical appraisal tool and provided feedback on the final manuscript.

Objective of the review

The objective of this systematic review was to evaluate the existing literature to assess the effectiveness of spawning habitat creation or enhancement for substrate-spawning fish.

Primary question

What is the effectiveness of spawning habitat creation or enhancement for substrate-spawning fish?

Components of the primary question

The primary study question can be broken down into the study components:

Subject (population):	substrate-spawning fish in temperate regions (covering a variety of substrate types as per Balon [4, 5]).
Intervention:	habitat creation or enhancement.
Comparator:	no intervention.
Outcomes:	use of habitat and the presence of eggs, survival/success of nests or eggs, presence of spawning adults.

Taylor *et al. Environ Evid* (2019) 8:19

Page 4 of 31

Secondary questions

The secondary questions are meant to help guide the overall goals of the systematic review and to ensure that areas of interest are encompassed in the methods. The secondary questions for this systematic review are:

1. Under what circumstances is spawning habitat creation or enhancement effective? In particular:

   a. To what extent does spawning habitat creation or enhancement effectiveness differ among interventions?
   b. What spawning habitat creation or enhancement measures are most effective for particular taxa?
   c. Is ecosystem type (lotic vs. lentic waterbodies) associated with intervention effectiveness?
   d. Is species life stage associated with intervention effectiveness?

2. Does the time since habitat creation or enhancement influence intervention effectiveness?

Methods

This review followed detailed methods described in the a priori systematic review protocol [33] and was performed according to the guidelines provided by the Collaboration for Environmental Evidence [22].

Search for articles

This systematic review was based on literature searches using five publication databases, one search engine, and 29 specialist websites (see Additional file 1). In a deviation from the protocol, the first 500 results from Google Scholar were used as opposed to the first 200 results and 29 websites were searched as opposed to 31 (see Additional file 1). Reference sections of accepted articles and 52 relevant reviews (see Additional file 2) were hand searched for any relevant titles that were not found using the search strategy. The majority of the literature searches were performed in December 2017–February 2018 (see Additional file 1). We issued a call for evidence to target sources of grey literature through relevant mailing lists (Canadian Conference for Fisheries Research, American Fisheries Society), and through social media (Twitter, Facebook) in February and November 2017. The Advisory Team also distributed the call for evidence to relevant networks and colleagues.

Estimating comprehensiveness of the search

To ensure the relevant articles were captured by the search, our search results were checked against a benchmark list of relevant papers provided by the Advisory Team (see Additional file 1). We also searched the reference lists of papers, as mentioned above, until the reviewer deemed that the number of relevant returns had significantly decreased. This increased the likelihood that relevant articles not captured by the literature search were still considered.

Article screening and study eligibility criteria

The literature found in publication databases and Google Scholar was screened for eligibility in EPPI Reviewer (eppi.ioe.ac.uk/eppireviewer4). Due to restrictions in exporting search results, the Waves database results were screened in a separate Excel spreadsheet. Prior to screening, duplicates were identified using a function of EPPI Reviewer and then were manually removed by one reviewer (JJT). One reviewer (James Monaghan [JM]) manually identified and removed any duplicates in the Waves spreadsheet.

Screening process

In a deviation from the protocol, the literature was screened at two distinct stages (1) title and abstract and (2) full-text, as opposed to three distinct stages. This change was made to allow for more efficiency by screening both the title and abstract at the same time. Prior to screening the full set of results, a consistency check was done at title and abstract where two reviewers (JM and Jill Brooks [JB]) screened 441/4419 articles (10% of the articles included in EPPI Reviewer; not including grey literature or other sources of literature, or the articles in the Waves spreadsheet). The reviewers agreed on 93.8% of the articles. A third reviewer (JJT) was consulted to resolve any disagreements between screeners and improve consistency before moving forward.

A consistency check was done again at full-text screening with 21/205 articles (10% of the articles included in EPPI Reviewer; not including grey literature or other sources of literature, or the articles in the Waves spreadsheet). The two reviewers (JM and JB) initially agreed on only 61.9% of articles but it was determined that the discrepancies were based largely on interpretation of the inclusion criteria for population (i.e. juveniles; as described below). After discussing disagreements with JJT and clarifying inclusion criteria, JM and JB agreed on 90.48% of articles and screening was allowed to continue. The remaining articles were split between JM and JB for screening. Reviewers did not screen studies (at title and abstract or full-text) for which they were an author.

Articles excluded based on full-text screening can be found in Additional file 2 along with their reason for exclusion.

Taylor *et al. Environ Evid* (2019) 8:19

Page 5 of 31

Eligibility criteria

Our eligibility criteria outlined below are based on the components of our primary question (population, intervention, comparator, and outcome).

Eligible populations Populations of substrate-spawning fish in north (23.5°N to 66.5°N) or south (23.5°S to 66.5°S) temperate regions were the subjects of this review. Spawning strategy included lithophils and phytophils as defined by the reproductive guilds described in [4]. Herein, substrate-spawning will include both substrate- and vegetation-spawning fish (e.g., northern pike as an example of a vegetation-spawning fish). The relevant subjects included all fish from egg (i.e., zygote or developing embryo) and larval stage (i.e., yolk sac larval stage) to age-0 (e.g., alevin, fry, young-of-the-year [YOY] that are no longer dependent on a yolk sac) as well as spawning adults. A decision was made by the Advisory Team to modify the criteria described in the protocol [33] to exclude articles focusing entirely on juvenile fish. This decision was made to ensure the focus of this review remained on effectiveness of spawning habitat and not nursery/rearing habitat. Therefore, any non-spawning fish described by authors as older than 1 year (e.g., age 1+, smolt) was excluded. One could argue that some age-0 fish may not be a using the habitat as spawning habitat, but rather using it as nursery habitat. However, we assumed because some researchers measured age-0 fish as the response to a spawning habitat creation or enhancement that this was a relevant and/or preferred age to measure the response. For instance, salmonids do not emerge from gravel redds until the yolk sac has been fully absorbed; hanging around redds for many weeks because of the protection provided by the spawning substrate. Once the yolk sac has been absorbed, fry relocate to more amenable nursery habitat. Researchers often set emergence traps on the spawning substrate to confirm successful salmonid egg development, and thus the traps would capture fry during emergence before relocation to nursery habitat (e.g., [34–36]). Also, a common technique for lake sturgeon (*Acipenser fulvescens*) spawning habitat assessment is to set larval drift nets immediately below spawning substrate to capture drifting fry seeking more amenable habitat than the spawning shoal, and confirm successful egg development (e.g., [37–39]). Therefore, we included age-0 (e.g., alevin, fry, YOY) fish when authors used this metric for evaluating spawning habitat creation and enhancements.

Eligible intervention Any creation or enhancement of spawning habitat was considered a relevant intervention. This included, but was not limited to, the addition of rock or plant material, creation of bays or artificial streams, modifications to the riparian zone, or addition of human-made structures. Based on discussions with the Advisory Team, interventions that involved flooding or altering flows were excluded, unless it was for the purpose of cleaning or altering the substrate by removing sediment (e.g., [40]). Allowing fish access to pre-existing habitat (e.g., adding a culvert) was not considered a relevant intervention for the purpose of this review as it does not involve creating or enhancing a habitat and was excluded.

Eligible comparator A non-intervention comparator was required in every included study. Study designs could take the form of Before/After (BA), Control/Impact (CI), Before/After/Control/Impact (BACI), or Randomized Control Trial (RCT). Relevant comparators included: (1) before data at the same study site, (2) a similar section of the same waterbody with no intervention applied, (3) a nearby waterbody with comparable habitat characteristics and no intervention applied. Contrary to what was proposed in the protocol [33], articles where evaluation of a spawning habitat creation or enhancement intervention was compared to an alternative level of that intervention (rather than to a no intervention comparator group) were excluded. We decided these studies were of limited value because they could not be compared to studies with non-intervention comparators in a quantitative analysis. Studies that reported only post-treatment monitoring data (i.e. no before or control site data) were excluded from this review. Simulation studies, review papers, and policy discussions were also excluded from this review.

Eligible outcomes Only direct outcomes in the form of a quantitative or qualitative measured effect of intervention were included. Relevant outcomes included, but were not limited to, abundance/density of nests, eggs, or age-0 fish, survival/success of nests or eggs, presence of spawning adults. Relative abundance estimates based on catch-per-unit-effort (CPUE) were also included, but indirect estimates using survival rate calculations or changes in physical habitat measures like spawning area were excluded.

Language Only English-language literature was included during the screening stage.

Study validity assessment
Study validity assessment

Critical appraisal of study validity was conducted on all studies included after full-text screening (Additional file 3). If a study contained more than one project (i.e., differed with respect to one or more components of critical appraisal; see Tables 1, 2), each project received an individual validity rating and was labelled in the data-extraction table with letters (e.g., "Avery 1996 A/B/C indicating that there are 3 projects

Table 1 Critical appraisal tool for study validity assessment

Category	Bias and generic data quality features	Specific data quality features	Validity	Design of assessed study
1	Selection bias: study design	Design (i.e., well-controlled)	High	BACI
			Medium	BA or CI
		Replication	High	Replication at level of intervention (n > 5)
			Medium	Replication at level of intervention (true n = 1 with pseudoreplication) or n = 2–5
			Low	Unreplicated (i.e. no variance, or variance across years only)
		Control matching	High	Control and treatment samples well-matched or likely well-matched at baseline
			Medium	Control and treatment samples moderately matched
			Low	Control and treatment samples poorly matched
2	Assessment bias: measurement of outcome	Measured outcome	High	Quantitative
			Medium	Quantitative approximations or semi-quantitative
			Low	Qualitative
3	Performance bias: baseline comparison	Other confounding environmental factors	High	Intervention and comparator sites homogenous
			Medium	Intervention and comparator sites moderately comparable with respect to confounding factors
			Low	Intervention and comparator sites hardly comparable with respect to confounding factors or lacking sufficient information to judge
				N/A if BA design and before measurement taken immediately prior to restoration

Reviewers provided a rating of high, medium, or low for each of the specific data quality features. Reviewers also had the opportunity to provide comments for each study based on external validity (generalizability)

Table 2 Terms related to study design and their definitions used throughout the systematic review

Term	Definitions
Article	An independent publication (i.e., the primary source of relevant information). Used throughout the review
Study	An experiment or observation that was undertaken over a specific time period at a particular site (i.e., ecologically independent sites from the same or different article). Used throughout the review
Project	Individual investigations within an independent study that differed with respect to ≥ 1 aspects of the study validity criteria (e.g., study design). Used in Review descriptive statistics and Narrative synthesis
Case	Situationally defined in text/visual aids. E.g., separate counts for different specific intervention comparisons (i.e., addition of sediment, gravel, boulders) within an independent study. Used in Review descriptive statistics and Narrative synthesis
Data set	(1) A single independent study from a single article; or (2) when a single independent study reported separate comparisons for different: (a) species, and/or (b) the same species but responses for different outcome subgroup categories (i.e., abundance, survival, body size), or different intervention subgroup categories (i.e., rock material, plant material, waterbody creation/extension, waterbody modification, human-made structures). The number of data sets was only considered for quantitative analyses

within the Avery 1996 article"). The critical appraisal framework (see Table 1) was developed based on features recommended by Bilotta et al. [41] and was adapted to incorporate components specific to the studies that answer our primary question. The framework used to assess study validity was reviewed by the Advisory Team to ensure that it accurately reflected the characteristics of an ideal study, regardless of resources or experimental/field restrictions. For example, in the case of habitat restoration, a high number of true replicates is not always feasible, due to spatial or financial constraints, and is therefore uncommon.

Taylor *et al. Environ Evid* (2019) 8:19

Page 7 of 31

After input from the Advisory Team, a sample size of n > 5 true replicates was deemed 'high' validity, while n = 2–5 was considered 'medium' and more feasible in the real world, and unreplicated was considered 'low' validity. Studies that provided only a count/number with no variance or presented a mean and variance across years and did not have replication within a year were considered unreplicated. See "Data-extraction considerations" below for details on pseudoreplication.

The criteria in our critical appraisal framework refer directly to internal validity (methodological quality), whereas external validity (study generalizability) was captured during screening or otherwise noted as a comment in the critical appraisal tool. The internal validity criteria included: study design (BA, CI, or BACI), replication (true or pseudoreplication), control matching [how well matched the intervention and comparator sites were at site selection and/or study initiation (e.g., physical characteristics)], measured outcome [quantitative, quantitative approximation (e.g., catch per unit effort, population estimates), semi-quantitative (e.g., absence before intervention and abundance data after intervention), or qualitative], and confounding factors [environmental or other factors that differ between intervention and comparator sites and/or times, that occur after site selection and/or study initiation (e.g., flood, drought, unplanned human alteration)]. Each criterion was scored as 'high' (low risk of bias), 'medium' (medium risk of bias), or 'low' (high risk of bias) based on the predefined framework outlined in Table 1. A study was given an overall 'low' validity if it scored low for one or more of the criteria. If the study did not score at least one low or all high for any of the criteria, it was assigned an overall 'medium' validity. Studies that scored only high for all of the criteria were assigned an overall 'high' validity. This approach assumes that equal weight was given to each criterion, which was carefully considered during the development of the predefined framework.

In most cases, study quality assessment and data extraction were performed simultaneously and by the same reviewer (JJT). If there was any uncertainty, another reviewer (TR) was brought into discuss and a consensus decision made. Initially, however, a consistency check was undertaken on 6/64 articles (9.4%) by JJT and TR. Meta-data and quality assessments on these studies were extracted by both reviewers, discrepancies were discussed and, when necessary, refinements to the meta-data extraction and quality assessment sheets were made to improve clarity on coding. Reviewers did not critically appraise studies for which they were an author.

Data coding and extraction strategy
General data-extraction strategy
Following full-text assessment, all included articles underwent meta-data extraction, regardless of their study validity category. Data extraction used a review-specific data-extraction form (Additional file 3). Extracted information followed the general structure of our PICO framework (Population, Intervention, Comparator, Outcome) and included characteristics such as: publication details, study location, study summary and timeline, population details, intervention and comparator details, and outcome variables. Abundance (including density, CPUE, and biomass), survival, and body size were treated as continuous outcome variables. Measures of abundance were used to address questions of broad differences in abundance, whereas survival and body size allowed understanding of the success and productivity of fisheries. For further syntheses, waterbody type was assessed as either (1) lotic (i.e., including: rivers, creeks, and sounds), or (2) lentic (including: lakes, wetlands, and reservoirs). During data extraction, redundant articles (i.e., articles that reported data that could also be found elsewhere or contained portions of information that could be used in combination with another more complete source) were identified and combined with the most comprehensive article (i.e., primary study source) (Additional file 4). Data on potential effect modifiers and other metadata were extracted from the primary study source or their supplementary articles.

In addition, all included articles underwent quantitative or qualitative data extraction. Sample sizes, outcome means (e.g., mean abundance of a fish species for the intervention and comparator groups) and measures of variability (e.g., standard deviation, standard error, confidence intervals of outcome means) were extracted if provided; data from figures were extracted using the data-extraction software WebPlotDigitizer [42] when necessary. If raw data, rather than means, were provided we calculated and recorded summary statistics ourselves. Where data or information were missing or unclear, we attempted to contact authors via email to retrieve the missing or unclear data.

Data-extraction considerations
There were a number of considerations made during data extraction (refer to Additional file 5 for a full summary of data-extraction considerations). For instance, first, if a single article reported data separately for sites we considered as ecologically independent (i.e. different interventions were applied to a number of sites, each with their own controls), we regarded these studies as independent

Taylor *et al. Environ Evid*   (2019) 8:19

Page 8 of 31

and assigned each study a separate "Site ID" (refer to Table 2 for term definitions).

A single independent study could also report separate relevant comparisons for: (1) different species, and/or (2) the same species but for different responses (i.e., abundance, survival, body size), or different interventions (i.e., rock material, plant material, waterbody creation/extension, waterbody modification, human-made structures). For quantitative synthesis, we treated these comparisons separately (i.e., separate rows in the database that share the same Site ID).

Replication within a study (i.e., group sample sizes) was considered at two levels: (1) independent intervention areas (i.e., separate waterbodies, or separate sections of a waterbody receiving treatment—true replicates), and (2) partly subsampled data, hereafter referred to as pseudoreplicated samples (i.e., in the sense that reported variances did not refer to the variability of true replicate means from (1) above but to the variability of subsamples within/across true replicates). For the former, we recorded the number of independent intervention areas as the level of true treatment replication. For the latter, we recorded the number of pseudoreplicated samples occurring, for example, at the plot or nest levels within an area (i.e., non-independent replicates). In cases of pseudoreplicated data (or presumed pseudoreplicated data), we made appropriate adjustments in the quantitative synthesis (see "Adjustment accounting for pseudoreplication"—Additional file 5).

Data-extraction consistency checking

As described above (see "Study validity assessment") in most cases, data extraction took place at the same time as the study quality assessment and by the same reviewer (JJT) after a consistency check was performed on a subset of the articles. Reviewers did not extract data from a study on which they were an author.

Potential effect modifiers and reasons for heterogeneity

For all 70 articles included on the basis of full-text assessment, the following data describing key sources of potential heterogeneity were extracted when available: waterbody type (e.g., creek, river, reservoir, or lake), fish taxa (at the family level), intervention type (i.e., rock material, plant material, waterbody creation, human-made structures, waterbody modifications, and any combination of these interventions; see Table 3 for intervention types and definitions), life stage [i.e., egg: nests, redds, or eggs (zygote or developing embryo); age-0: alevin, fry, YOY: adult spawners], and time since intervention. We consulted both the Advisory Team and similar published analyses [17] when selecting potential

Table 3 Intervention types assessed in this review along with definitions and codes

Intervention type	Code	Definition	Intervention category	Definition
Sediment	SED	Sediment addition	Rock material	Aggregate/rock (additions/alterations)
Gravel	GRAV	Gravel Addition		
Cobble	ROCK	Addition of cobble, boulders, small rocks, artificial reefs made of rocks		
Sediment trap	TRAP	Removal of sediment using a sediment trap		
Gravel washing	WASH	Gravel washing using pressure washers, or mixing substrate and allowing fine sediment to wash downstream, includes substratum raking, flooding to clean sediment		
Excavate	EXCAV	Excavation	Waterbody modification	Waterbody modifications (physical alterations)
Riparian	RIP	Riparian modifications, grading of banks		
Nutrient	NUTR	Addition of a nutrient		
Bay	BAY	Creation of bays	Waterbody creation	Creation of a new waterbody or extension of an existing waterbody
Stream	STREAM	Artificial stream, creek, spawning channels		
Structure	STRUC	Human-made structure (e.g., mesh tubes, PVC, masonry blocks, ceramic tiles)	Human-made structures	Human-made structures (additions)
Log	LOG	Addition of logs, large woody debris	Plant material	Plant material
Brush	BRUSH	Addition of brush bundles/straw bails		
Vegetation	VEG	Planting of vegetation including macrophytes		

Intervention types were assigned based on intervention details provided by authors. Intervention categories were assigned to combinations of one or more similar intervention types (i.e., Intervention type) and used in the Narrative and Quantitative Syntheses (i.e., to increase sample sizes of intervention type categories for meta-analyses). Intervention codes were used in forest plots for meta-analyses (i.e., visualize aids that plot mean effect sizes and 95% confidence intervals from individual comparisons)

Taylor *et al. Environ Evid* (2019) 8:19

Page 9 of 31

effect modifiers. After consultation with the Advisory Team, there were effect modifiers that were originally identified in our protocol that were removed from data extraction for this review. Due to limitations in time and resources, we did not search external to the article for geographic coordinates, climate region, substrate type, and spawning strategy as they were deemed to not be key sources of potential heterogeneity and/or were rarely reported within the primary articles. When data were sufficient and sample size allowed, these potential modifiers were used in meta-analyses (see "Meta-analyses" section below) to account for differences among data sets via meta-regression (see Table 2 for definitions of terms such as data set).

Data synthesis and presentation
Descriptive statistics and a narrative synthesis
Following full-text assessments, we included all relevant studies in an MS-Excel database (Additional file 3). Metadata on each study were used to generate descriptive statistics and a narrative synthesis of the evidence, including figures and tables.

Meta-analyses
Eligibility for meta-analysis Despite inclusion in the database, some studies were considered unsuitable for meta-analysis (and were not included in the quantitative synthesis). These were studies that: (1) were critically appraised as having low study validity (see Table 1); (2) did not report measures of outcome variability and/or data on sample sizes and these data could not be otherwise calculated; and (3) averaged across sampling years, the most recent Before and/or After years could not be isolated (i.e., not comparable with other studies).

Initial data preparation Prior to quantitative synthesis, BACI outcomes were converted to CI by subtracting data sampled before the intervention (B) from those sampled after the intervention (A) for each C and I site [i.e., C: (A-B) and I: (A-B); then means and variances were obtained by averaging across sites within each group] (see calculations in Additional file 6). Measures of variability were converted to standard deviations, if not reported as such (e.g., standard errors or confidence intervals).

Effect size calculation Because outcomes (e.g., abundance, CPUE, density, survival, body size) were not always reported in comparable units, we used the standardized mean difference (Hedges' *g*) as our effect size measure instead of raw mean differences. Hedges' *g* was calculated using the following steps [43], as shown below. Beginning with Cohen's d, the standardized mean difference was used to account for differences in the scale of measure-

ment across studies by dividing the mean difference in each study (i.e., the difference between the mean response to an intervention and the mean response to no intervention) divided by that study's pooled standard deviation:

$$d = \frac{\bar{X}_{G2} - \bar{X}_{G1}}{S_{\text{pooled}}} \tag{1}$$

where $\bar{X}_{G1}$ and $\bar{X}_{G2}$ were the means of group 1 (G1 = comparator group) and group 2 (G2 = intervention group). S_{pooled} was the pooled standard deviation of the two groups:

$$S_{\text{pooled}} = \sqrt{\frac{(n_{G2} - 1)S_{G2}^2 + (n_{G1} - 1)S_{G1}^2}{n_{G1} + n_{G2} - 2}} \tag{2}$$

where S = standard deviation, and n_{G1} and n_{G2} were the sample sizes of group 1 and group 2. The variance for *d* is given by:

$$V_d = \frac{n_{G1} + n_{G2}}{n_{G1} n_{G2}} + \frac{d^2}{2(n_{G1} + n_{G2})} \tag{3}$$

To convert from Cohen's *d* to Hedges' *g*, we used a correction factor that removes small sample size bias:

$$J = \left[1 - \frac{3}{4(n_{G1} + n_{G2} - 2) - 1} \right] \tag{4}$$

Then Hedges' *g* and associated variance (V_g) were calculated as:

$$Hedges'g = J \times d \tag{5}$$

$$V_g = J^2 \times V_d \tag{6}$$

Thus, a positive Hedges' *g* indicates that the response outcome (abundance, survival, or body size) was higher/longer in the created or enhanced spawning habitat areas than in areas with no intervention.

Quantitative synthesis All meta-analyses were conducted in R 3.4.3 [44] using the *rma.mv* function in the metafor package [45].

To determine whether habitat creation or enhancement measures improve, on average, substrate spawning fish responses compared to controls, we first conducted random-effects meta-analyses using restricted maximum-likelihood (REML) to compute weighted summary effect sizes for each outcome separately (i.e., abundance, survival, and body size). To further account for multiple study comparisons within a study site and species outcomes being reported from the same site (see "Combining data across outcomes or multiple comparisons within a study" in Additional file 5 for full adjustment

summary), Site ID was included as a random factor in each model. The summary effect size was considered to be significantly different from zero (i.e. there was a significant either positive or negative effect of intervention) when the 95% confidence intervals (CI) did not overlap zero. Heterogeneity in effects was calculated using the Q statistic, which was compared against the χ^2 distribution, to test whether the total variation in observed effect sizes (Q_T) was significantly greater than that expected from sampling error (Q_E) [46]. A statistically significant Q indicates greater heterogeneity in effect sizes (i.e., individual effect sizes do not estimate a common population mean), suggesting there are differences among effect sizes that have some cause other than sampling error. We also produced forest plots to visualize mean effect sizes and 95% confidence intervals from individual comparisons. The purpose of these summary effect sizes was to identify general trends in the evidence base. It is important to note, that a lack of significance does not indicate no significant patterns within the evidence base. Furthermore, a lack of significance can only be interpreted as a lack of evidence for an effect if there is no indication of heterogeneity. Additionally, if a significant pattern is detected within the evidence base, interpretation of summary effects should include some consideration for context. For example, if a significant positive summary effect of habitat creation or enhancement on fish abundance is detected, it should not necessarily be interpreted as evidence that any habitat creation or enhancement measure designed for substrate spawning fish will improve fish abundance.

Given that Hedges' *g* may not be easily interpretable, we attempted to convert *g* to a weighted-mean percent change in intervention effectiveness by plotting the relationship between *g* and the percent change in intervention effectiveness:

$$\frac{\bar{X}_{G2} - (\bar{X}_{G1} + q)}{\bar{X}_{G1} + q} * 100, \tag{7}$$

where $\bar{X}_{G1}$ and $\bar{X}_{G2}$ were the means of group 1 (G1 = comparator group) and group 2 (G2 = intervention group). Since percent change cannot be computed when $\bar{X}_{G1} = 0$, we added a small constant $q = 0.01$ to $\bar{X}_{G1}$ for each data set. Also, because the calculation for percent change has no upper or lower bound, it can be excessively large when the comparator mean is small; to address this, we trim these extreme values by bounding the percent change to ± 100. For all analyses, we accompany weighted-mean effect sizes with weighted-mean percent changes and 95% confidence intervals from individual comparisons (see Additional file 6). It should be noted that while this metric provides an indication

of the relative proportional change in the intervention effectiveness, it is unclear, in a broad sense, how closely related this metric is to Hedges' *g*. As such, this metric should only be used to aid with interpretation of effect size estimates.

Despite our effort to reduce publication bias by including data available in grey literature, the results could still be flawed if there was a bias towards publishing only positive or statistically significant results. Therefore, we examined the robustness of our models by testing for publication biases in two ways. First, we used visual assessments of funnel plots (i.e., scatter plots of the effect sizes of the included studies versus a measure of their precision e.g., sample size, standard error, or sampling variance) [47]. If no bias is present, the funnel plot should be funnel-shaped, with a wider spread of effect sizes for less precise (smaller) studies and decreasing spread as study precision increases (larger studies). We produced funnel plots using 1/square root of sample size, since standard errors have been shown to be inappropriate for funnel plots of standardized effect sizes [48]. In these plots, as study sample size increases (1/sqrt(k) decreases) we should expect the variance in the effect size to decrease if no bias is present. Second, in an attempt to judge the robustness of results against publication bias, the fail-safe numbers were calculated using the method as described by Rosenberg [49] specified with the *fsn* function in the metafor R package [45]. A fail-safe number estimates the number of non-significant unpublished studies required to eliminate a significant (weighted) overall effect size [49, 50]. The fail-safe number is often considered robust if it is greater than 5k + 10, where k is the number of effect sizes in the analysis (see [51]).

To test for associations between effect size and moderators in relation to our secondary research questions, we used mixed-effects models for categorical moderators (i.e., intervention type, ecosystem type, and life stage) and meta-regression for continuous moderators (i.e., time since intervention), estimating heterogeneity using REML. Because studies did not always report information for all of our moderators of interest to combine them in a single model (nor did sample size allow for this—see below), we first conducted random-effects models (unmoderated models) using a subset of responses (e.g., a subset of abundance effect sizes) that maximized the number of effect sizes for testing the influence of the moderator variable in question. Then using this same subset, we conducted a mixed-effects model/meta-regression including the moderator of interest. To further account for multiple study comparisons within a study site and species outcomes being reported from the same site, Site ID was included as a random variable in each model. We restricted the number of fitted parameters (*j*)

Taylor *et al. Environ Evid* (2019) 8:19

Page 11 of 31

in any model such that the ratio *k/j*, where *k is* the number of effect sizes, was greater than 5, which is sufficient in principle to ensure reasonable model stability and sufficient precision of coefficients [52]. The small number of effect sizes did not permit the construction of models with multiple variables, therefore weighted simple mixed-effects models were used throughout analyses.

We only performed analyses of categorical moderators where there were sufficient combinable data sets (i.e., > 2 data sets from ≥ 2 independent studies) for each moderator category. Thus, in some cases, we either combined similar categories to increase the sample size (see Table 3 Intervention Categories and detailed in "Results" below) or deleted the categories that did not meet the sample size criteria. For example, the intervention category 'Rock material' included the addition or alteration of gravel, cobble, or sediment.

For all data analyses, total heterogeneity, Q_T, was partitioned into heterogeneity explained by the model, Q_M, and heterogeneity not explained by the model, Q_E (i.e. $Q_T = Q_M + Q_E$). The statistical significance of Q_M and Q_E were tested against a χ^2 distribution. Due to skewness of the data, time since intervention (continuous moderator) was log-transformed before analysis.

Results
Review descriptive statistics
Literature searches and screening
A search of 5 databases and Google Scholar returned 5164 individual records (see Additional file 1), which resulted in 4611 articles after duplicate removal. Of those remaining articles, 4363 were removed after screening at title and abstract, leaving 244 potentially relevant articles. We were able to obtain all but one full text (see Additional file 3), leaving 243 articles to screen at full-text. The majority of articles were excluded at full text because of an irrelevant intervention (i.e., not creation/enhancement of a spawning habitat), population (i.e., study reported data for fish > age-0 or non-temperate or non-substrate spawning fish), or outcome (i.e., irrelevant outcomes or lack of results). All articles excluded at full text along with reasons for their exclusion can be found in Additional file 2. From the databases and Google Scholar, 46 articles were included at full-text.

Searching the bibliographies of articles identified as relevant at either title and abstract or full-text review stage resulted in an additional 17 articles included at full-text. Website searches and grey literature solicitation provided an additional five articles included at full text. A total of 70 articles were deemed relevant at full-text, six of which were considered supplemental (redundant) articles because they overlapped with other included articles providing only additional

information (e.g., extra years of data, intervention information). A total of 75 studies from 64 articles (see definitions in Table 2) were included in this systematic review and underwent data extraction and critical appraisal (Fig. 1).

Study validity assessment
The majority of projects (60.0%) were assigned an overall 'low' study validity, whereas 40.0% were assigned an overall 'medium' validity, and 0 studies were of overall 'high' validity. BA or CI study designs were used in 86.3% of projects, which resulted in a medium study validity for the study design category (see Table 4). Of the 14 projects that used a BACI study design, none had high replication at the level of the intervention (n > 5), which led to no studies with high validity in this category. Among the projects that received an overall low study validity, most (84.2%) lacked replication. This included projects that provided: (1) only a number or count, with no mean or variance; (2) a mean and variance across years; or (3) a mean with no variance and no raw data. The majority of the BACI study designs were either pseudoreplicated or unreplicated, and only two projects had 2–4 replicates. Most projects scored high validity (low risk of bias) in the categories of control matching, measured outcome, intervention, and confounding factors.
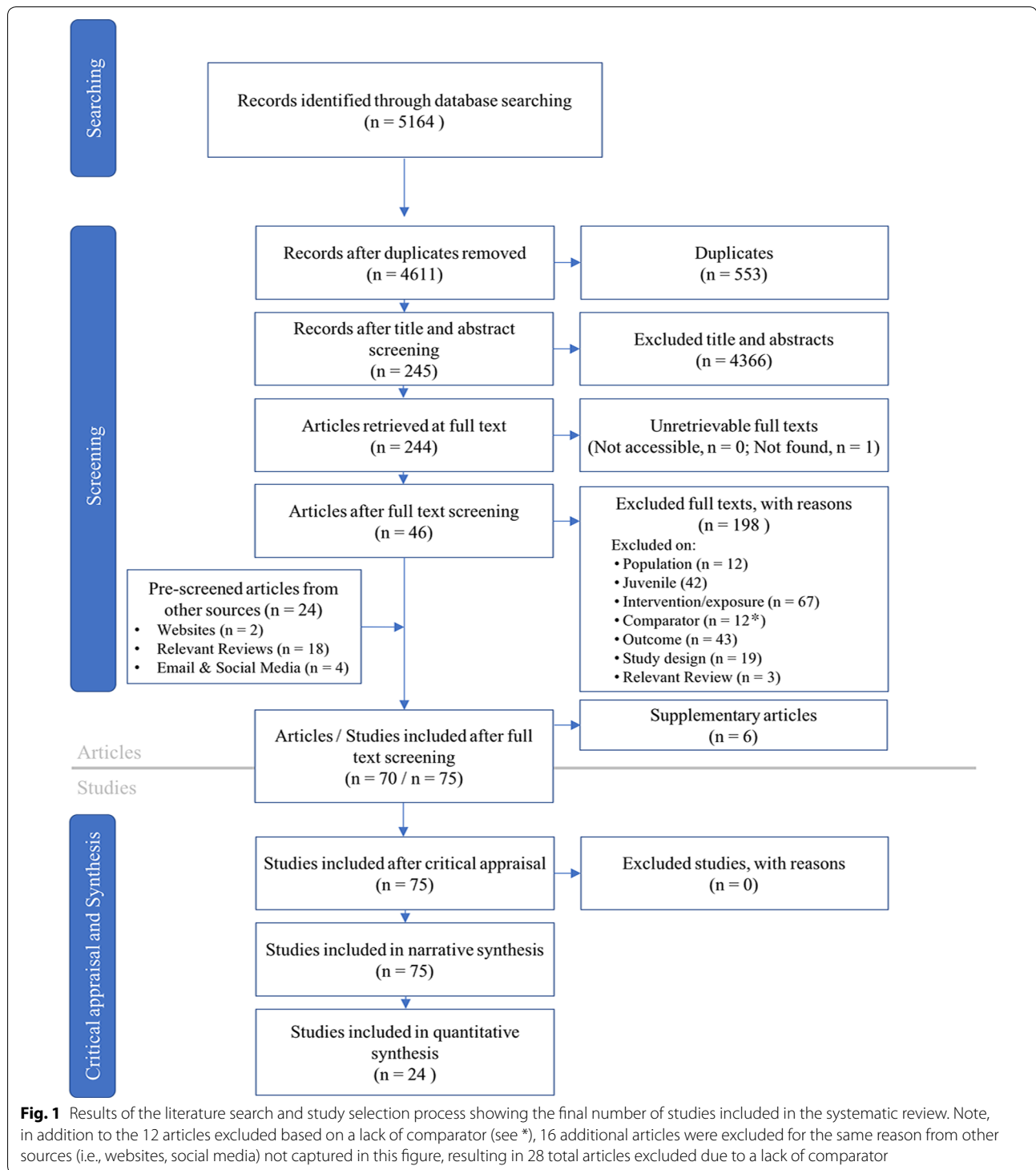
Publication year
Included articles were published from 1962 to 2016, with the number of publications increasing over time. From 1962 to 1990, grey literature made up a larger proportion of the total articles than in more recent years (Fig. 2). Critical appraisal of all studies indicated that study validity tended to improve over time (Fig. 3), with medium validity studies making up 56%, 49%, and 65% of studies in each of the last three decades (1991–2016) compared to 29%, 29%, and 0% in the previous three decades (1961–1990).

Narrative synthesis
The narrative synthesis was based on all 75 studies from 64 articles, regardless of study validity. A database of these studies with descriptive meta-data, coding and quantitative data is available in Additional file 3.

Study location
The vast majority of studies included in this systematic review took place in North America (78.1%), primarily in the United States of America (49.3%) (Fig. 4). European countries made up 19.2% of all studies with 1 study each in Sweden, Switzerland, and Finland, 2 studies each in Norway and England, 3 studies in Germany, and 4 studies

Fig. 1 Results of the literature search and study selection process showing the final number of studies included in the systematic review. Note, in addition to the 12 articles excluded based on a lack of comparator (see *), 16 additional articles were excluded for the same reason from other sources (i.e., websites, social media) not captured in this figure, resulting in 28 total articles excluded due to a lack of comparator

in Denmark. Most interventions were applied in creeks or rivers (80% of studies), with a few in lakes/ponds (9.3%), reservoirs (9.3%) or in a sound (1.3%).
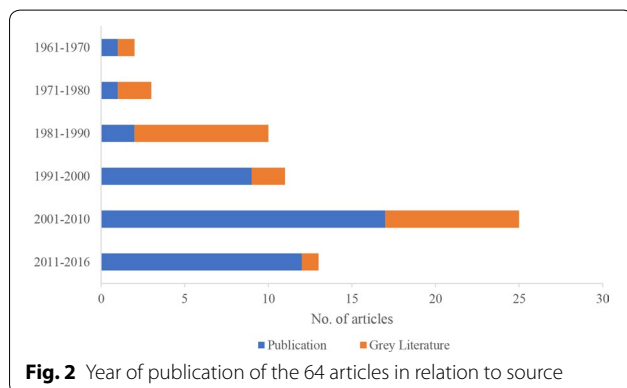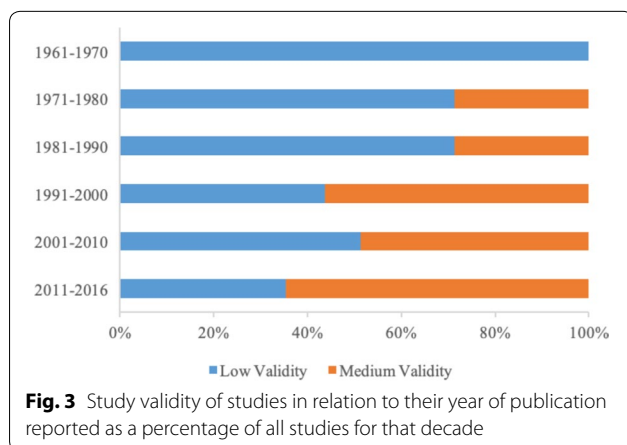
Study design

Of the 75 studies included in the systematic review, 25 implemented a Before/After (BA) design. Other than one BA study [53] that had true intervention replication, most BA studies were either unreplicated or pseudoreplicated.

Table 4 Results of study validity assessment using the critical appraisal tool (see Table 1)

	High			Medium			Low		
Study design	BACI		13	BA or CI		82	N/A		–
Replication at level of intervention	n > 5		5	n = 1–4		42	Unreplicated		48
Control matching	Well matched at site selection/study initiation		80	Moderately matched at site selection/study initiation		9	Poorly matched at site selection/ study initiation		6
Measured outcome	Quantitative		79	Semi-quantitative or quantitative approximation		13	Qualitative		3
Confounding factors	Intervention and comparator sites are homogenous during study		74	Intervention and comparator sites moderately comparable with respect to confounding factors during study		10	Intervention and comparator sites hardly comparable with respect to confounding factors during study or lacking information to judge		11

Numbers indicates the number of projects that received the critical appraisal score for each criterion

Fig. 2 Year of publication of the 64 articles in relation to source

Fig. 3 Study validity of studies in relation to their year of publication reported as a percentage of all studies for that decade

A Control/Impact (CI) design was used in 42 of the included studies (Table 4). These studies were more likely to include true replication, which often took the form of multiple control/impact streams or study sites. A Before/After/Control/Impact (BACI) design was used in 13 of the studies (Table 4), and included a minimum of 1 year of before and after data and 1 control and 1 impact site.
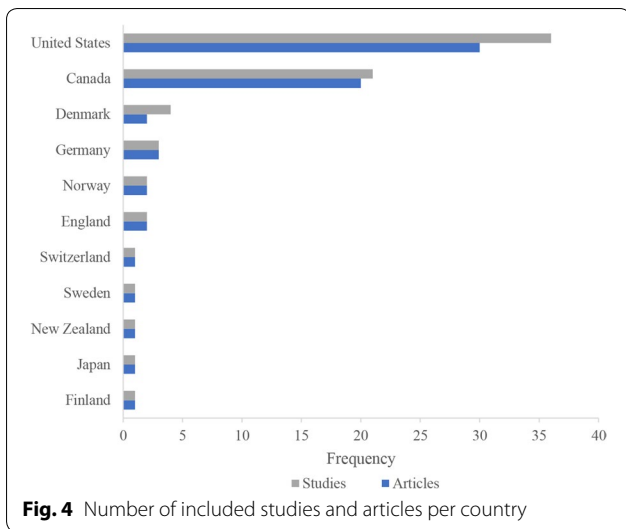
The number of studies listed above exceeds the total number of studies because some studies included more than one project. A number of studies (19) were excluded from this review at full-text screening based on lack of comparator ('post-treatment' design). These studies often described a habitat creation or enhancement with monitoring data, but no data from before the intervention or at a control site.

Population
Included studies investigated the effect of spawning habitat creation or enhancement on 40 species from 25 genera and 11 families (Fig. 5). Salmonidae were the most common family studied (63.7% of studies) including *Coregonus* (1 study), *Oncorhynchus* (21), *Salmo* (21), *Salvelinus* (13), and *Thymallus* (2). Four studies grouped several species across genera or families and were therefore not included in Fig. 5. The most frequently studied species were brown trout (*Salmo trutta*; 16 studies), brook trout (*Salvelinus fontinalis;* 9), walleye (*Sander vitreum*; 7), coho salmon (*Oncorhynchus kisutch*) and chinook salmon with 5 studies each, and chum salmon (*Oncorhynchus keta*), rainbow trout (*Oncorhynchus mykiss*), sockeye salmon (*Oncorhynchus nerka*), and lake trout (*Salvelinus namaycush*) with 4 studies each.

Interventions
The majority of studies applied only one intervention type to create or enhance spawning habitat for substrate-spawning fish (72.9% of cases, Table 5). In the remaining studies (29), an intervention was applied in combination with one or more different interventions. The number of cases exceeds the number of studies, because a study was considered to contain multiple cases if it investigated the effect of different interventions within a study, or if different critical appraisal scores were assigned to an intervention based on experimental design.

Taylor *et al. Environ Evid*   (2019) 8:19

Page 14 of 31

Fig. 4 Number of included studies and articles per country

Manipulation of rock material (see Table 3 for examples and definitions) was the most common intervention used across studies that involved the application of only one intervention (43.9% of cases, Table 5). Gravel (e.g.,

addition of spawning beds) and cobble (e.g., rock piles or artificial reefs) were the most common interventions applied, followed by gravel washing (e.g., pressure washing). Very few studies tested the removal of sediment through the installation of a sediment trap or the addition of sediment. Waterbody creation was used alone as an intervention in 15.0% of cases, including the creation of a bay or an artificial stream. The addition of plant material was less frequently applied as an individual intervention (e.g., brush, logs; 6.5% of cases), as was the addition of human-made structures (e.g., masonry blocks, ceramic tiles; 4.7% of cases), and waterbody modifications (e.g., grading of banks; 2.8% of cases).

In 27.1% of all cases, an intervention was applied in combination with one or more different interventions (Table 5). Of those cases, a combination of 2 interventions (e.g., cobble and gravel, log and human-made structure) was used in 17 cases. A combination of 3 interventions was used in 11 cases (e.g., cobble, gravel, and log), and only 1 case described a combination of 4 interventions (e.g., cobble, gravel, human-made structure, and log). Full definitions of intervention categories can be found in Table 3.

Fig. 5 The number of studies per family and genus. The number of studies per family is shown in brackets adjacent to the family name. The number of studies shown exceeds the total number of included studies because data for multiple genera were often presented within a study
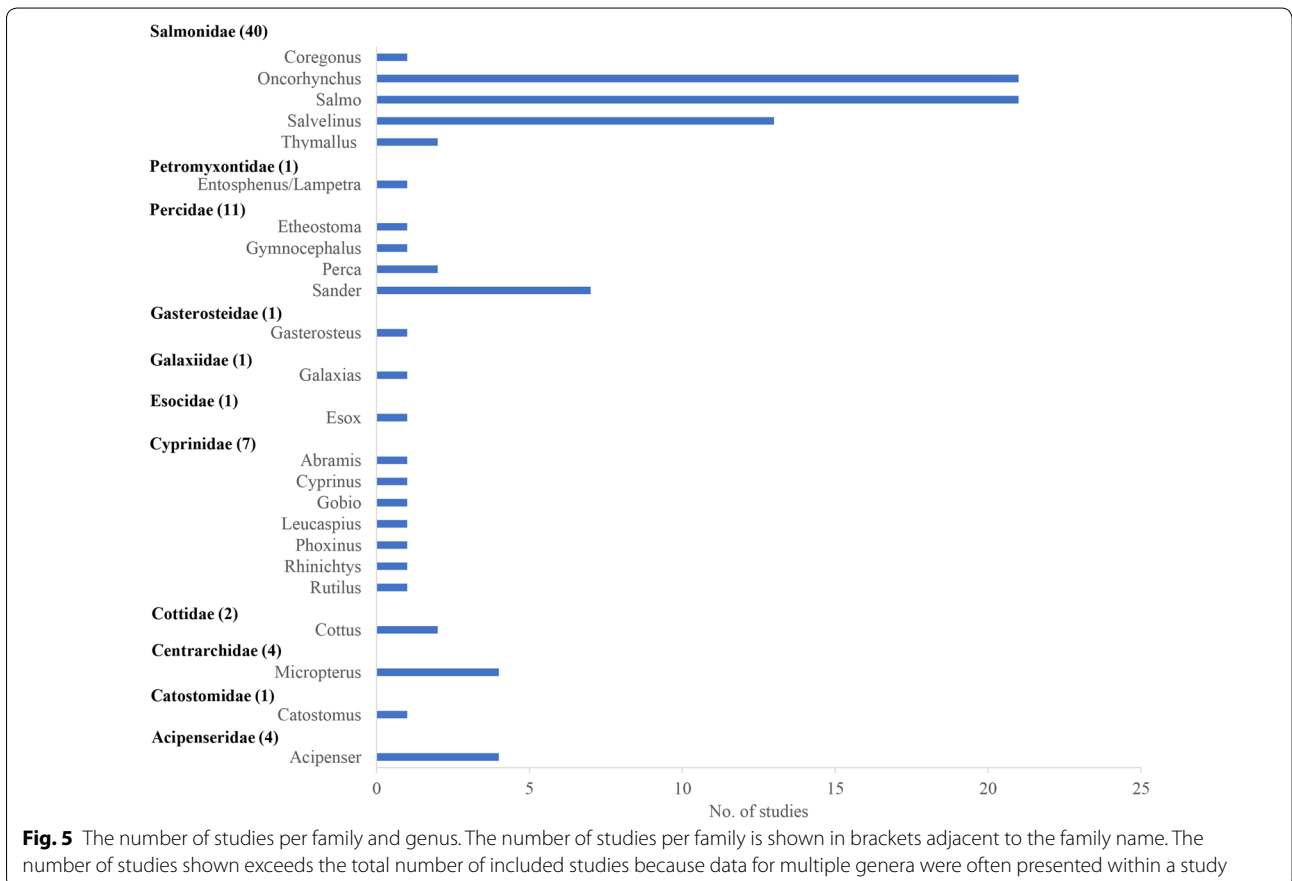
Table 5 Number of cases with low or medium study validity for each intervention type either alone or in combination

	Study validity	
	Low	Medium
Alone		
Rock material		
Gravel	6	4
Cobble	14	9
Sediment	0	1
Sediment trap	1	1
Gravel washing	4	7
Plant material		
Brush	4	0
Log	1	2
Waterbody creation		
Bay	0	2
Stream	9	5
Human-made structure		
Structure	3	2
Waterbody modification		
Excavate	1	1
Riparian	1	0
Combinations		
Log + structure	0	1
Log + excavate	0	1
Riparian + vegetation	0	1
Cobble + sediment trap + gravel	1	0
Gravel + structure	1	2
Cobble + structure	0	2
Cobble + log	2	2
Cobble + nutrient + log	0	1
Cobble + gravel	3	2
Cobble + gravel + excavate	0	1
Cobble + gravel + log	1	0
Cobble + gravel + structure + log	1	0
Sediment trap + gravel + log	3	3
Log + gravel + structure	0	1

Studies involving Salmonidae, Percidae, Acipenseridae, and Catostomidae most often applied the manipulation of rock material alone as an intervention to enhance spawning habitat (24, 6, 3, 1 studies respectively; Fig. 6). Studies involving Petromyzontidae, Galaxiidae, Cottidae, and Centrarchidae most often applied the addition of plant material alone (one, one, one, and two studies respectively). Studies involving Gasterosteidae and Esocidae (one study each) used a waterbody creation as an intervention, while for Cyprinidae (two studies) a combination of interventions was used to enhance spawning habitat.

Measured outcomes

The vast majority of studies reported a metric of abundance (including abundance, density, CPUE, and biomass; 91 studies) as an outcome, whereas 12 studies reported a survival metric, and 6 studies reported body size metrics (Fig. 7). The studies focused mostly on early life-stage outcomes as opposed to spawning adults. Studies reporting egg data (including nests/redds, zygote or developing embryo, and larvae) and age-0 data (including alevin, fry, YOY) were in almost equal proportion (47 and 51 studies, respectively). Spawning adults were reported in 11 studies, all of which reported abundance metrics. Most studies (63) reported outcomes as quantitative data, whereas 12 studies reported a quantitative approximation (e.g., CPUE, population estimates), 5 studies reported semi-quantitative data (e.g., presence before intervention and quantitative values after intervention), and 2 studies reported only qualitative outcomes. Nearly equal numbers of studies did (38 studies) or did not (41 studies) present intermediate time points of data (i.e., more than just one after year). Most often, studies reported data collected 1 or 2 years post-intervention (44 and 55 cases, respectively). Very few studies reported long-term monitoring (Fig. 8), with cases containing data from over 8 years (96 months) post-treatment stemming from only two articles [54, 55]. Studies that did not provide dates were not included in Fig. 8.

Quantitative synthesis
Description of the data

Of the 75 studies (from 64 articles) included in the narrative synthesis, 22 studies (from 20 articles) with 53 data sets were included in the quantitative synthesis. We excluded studies for the following reasons: (1) studies were evaluated as having low study validity (43 studies); (2) measures of outcome variability and/or data on sample sizes were not reported or could not be calculated (7 studies); (3) data were averaged across sampling years, not allowing the most recent before and/or after years to be isolated (1 study) (see details of these studies in Additional file 6). Additionally, we excluded 2 further studies because of differences in the comparators used which resulted in different interpretation of effect size estimates (i.e., they were not comparable to other effect sizes). For both BA study designs, and CI designs that compared control and impact sites from the same stream (i.e., impact sites were sites within the same stream where the intervention was applied but otherwise, control and impact sites were similar), based on the Hedges' *g*, we would expect a positive estimate if the outcome (abundance, survival, or body size) was higher/longer in the created or enhanced spawning habitat areas (or the after intervention time period) than in areas with
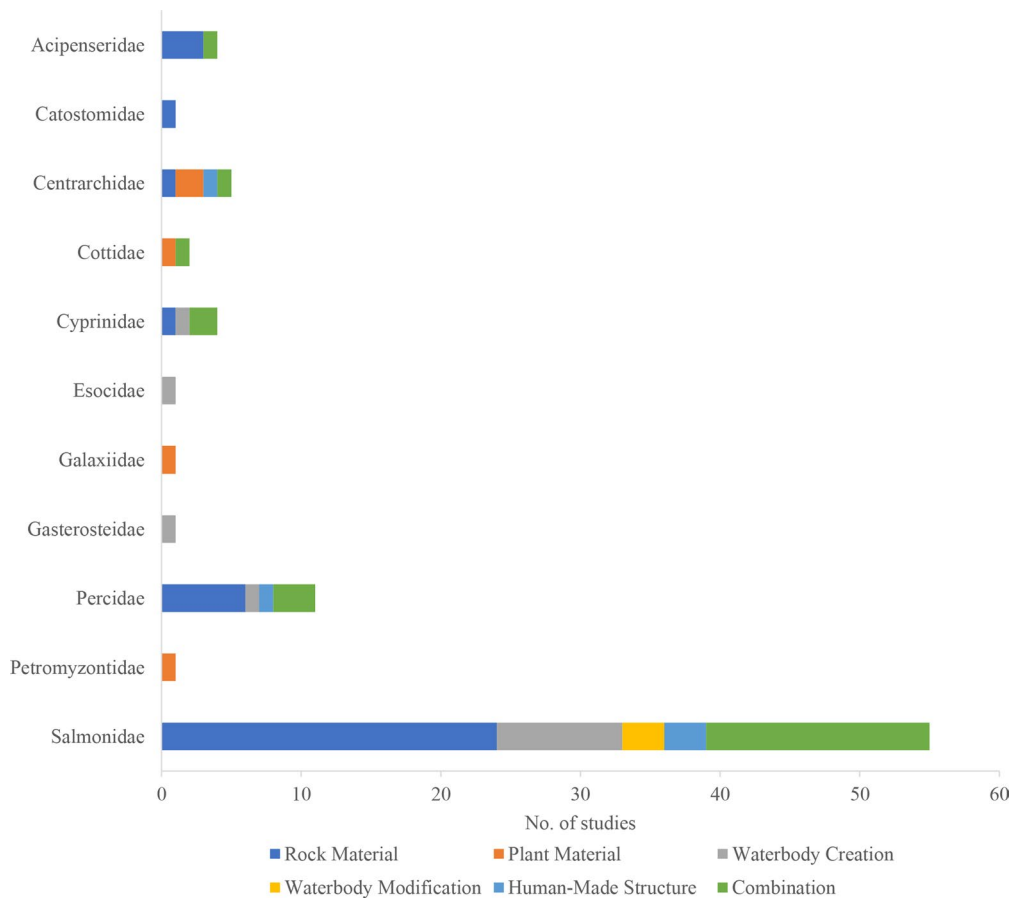
Fig. 6 The number of studies per family in relation to the intervention type applied. Combination refers to any number of interventions applied simultaneously. The number of studies shown exceeds the total number of included studies because data for multiple families or intervention categories were often presented within a study
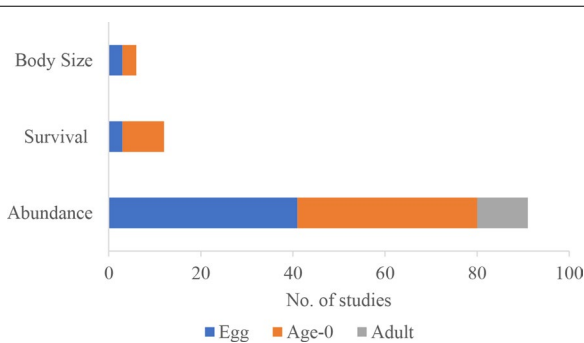

Fig. 7 The number of studies per outcome metric in relation to the life stage presented. The number of studies shown exceeds the total number of included studies because data for multiple life stages were often presented within a study. Egg: nests, redds, or eggs (zygote or developing embryo); Age-0: alevin, fry, YOY; Adult: adult spawners

no intervention (or the before intervention time period). However, for CI designs that compared impact sites that were degraded and to which an intervention was applied

with control sites that represented a more natural condition, based on Hedges' *g*, we would expect: (1) a positive effect size (i.e., $g > 0$) if the intervention resulted in a larger improvement than the control (natural condition), or (2) a neutral effect size (i.e., $g = 0$) if the outcome at the impact sites was similar to the outcome at the control sites. Because comparator types were not comparable across all study designs, and since there were too few effect sizes to subgroup comparator types, we excluded these two studies from further analyses (i.e., [56, 57]). All 22 studies included in the quantitative synthesis were assessed as having 'Medium' study validity.

Data sets included in the quantitative synthesis were predominantly from North America (Canada, 13; USA, 17), followed by some from Europe (22), and single study from Asia. The majority of data sets were from studies conducted in lotic ecosystems (91% of data sets), including rivers (83%), creeks (4%), and sounds (4%), and a few data sets were from lentic systems (9%), including reservoirs (6%), and lakes (4%). Eighty-three percent
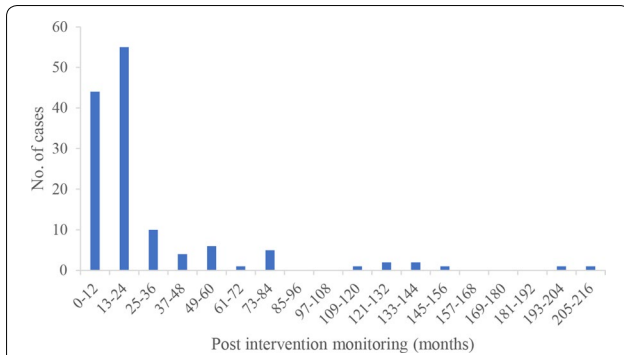
Fig. 8 The number of cases in relation to the number of months between the last intervention and the last reported sampling time. In this context, a study is made up of multiple cases if it includes more than one life stage or intervention per species

of the data sets used a CI study design, whereas 6% and 11% used a BA or BACI design, respectively (Additional file 6).

Among the 53 data sets extracted, 39 data sets in 17 studies reported species (90% of data sets) or group (i.e., species across genera or families; 10% of data sets) abundances (i.e., counts (15), CPUE (15), density (8), and biomass (1), see Additional file 6). Six data sets in 6 studies reported species survival, and 8 data sets in 4 studies reported species (88% of data sets) or group (12%) body size outcomes.

The majority of the data sets implemented a single intervention type (33/53 data sets). Of the data sets that implemented single interventions, most were evaluations of the effectiveness of the creation of a new waterbody (i.e., stream) or extension of an existing waterbody (i.e., bay), and the addition or alteration of rock material

(including sediment, gravel, cobble, boulders, and/or gravel washing or substrate raking) (Table 6). The addition of plant materials (i.e., logs) was used less frequently (Table 6; Additional file 6).

There were relatively fewer data sets that used a combination of intervention types to create or enhance spawning habitat (22/53 of data sets). The most common combination of intervention types included the addition of plant material (i.e., logs or vegetation) with: (1) physical alteration to the waterbody (i.e., riparian modifications or excavation), and (2) the addition of rock material (i.e., boulders/cobble) (Table 6; Additional file 6).

Among the 53 data sets, 26 fish species from 18 genera and 9 families were targeted for spawning habitat restoration (Table 7). The most commonly targeted species were from the Salmonidae family and included brown (6 data sets) and chinook salmon (3).

Data sets reporting outcomes using age-0 (fry to age-0) as the life stage made up the largest portion (68% of data sets), whereas eggs (including nests/redds, eggs and larvae) made up 30% of data sets. Only a single data set collected outcomes using adult spawners (along with redd counts) (Table 6; Additional file 6).

Information on the time since habitat creation or enhancement (time between the last intervention and the last outcome measure) was reported in 38/53 of the data sets. Most data sets were short-term evaluations of spawning habitat restoration, with 47% of the available data sets reporting restoration evaluations between 12 and 24 months after the last intervention was implemented, and 18% of data sets reporting evaluations less than 12 months after the most recent restoration measure was applied. We found some data sets reporting

Table 6 Sample sizes (number of data sets) for the three different outcome metrics by life stage and interventions

	Abundance				Survival				Length			
	Total	Egg	Age 0	Adult	Total	Egg	Age 0	Adult	Total	Egg	Age 0	Adult
Alone												
Rock material	6	5 (1)	0	(1)	4	3	1	0	2	2	0	0
Plant material	4	2	2	0	0	0	0	0	3	1	2	0
Waterbody creation	14	2	12	0	0	0	0	0	0	0	0	0
Combinations												
Rock material + human-made structures	0	0	0	0	1	1	0	0	0	0	0	0
Rock material + plant material	4	0	4	0	0	0	0	0	3	0	3	0
Rock material + plant material + human-made structures	1	0	1	0	0	0	0	0	0	0	0	0
Rock material + plant material + waterbody modifications	2	0	2	0	0	0	0	0	0	0	0	0
Plant material + human-made structures	1	1	0	0	1	1	0	0	0	0	0	0
Plant material + waterbody modifications	7	0	7	0	0	0	0	0	0	0	0	0
Total	39				6				8			

Numbers in brackets indicate a data set that included > 1 life stage

Table 7 Sample sizes (number of data sets) for the three different outcomes by interventions

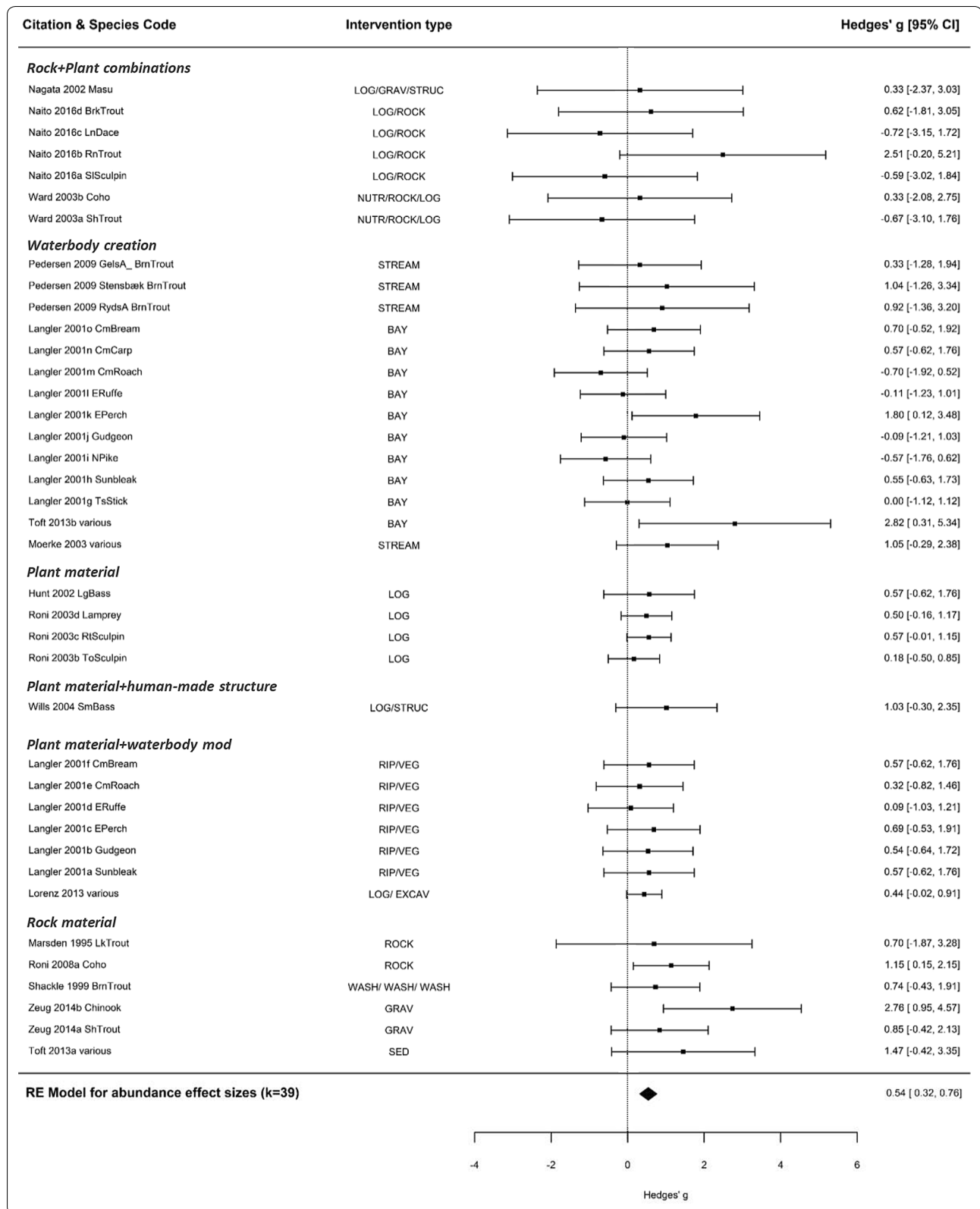|	Abundance	Survival	Body size
Total	37	6	8
Acipenseridae	0	0	1
Rock material	0	0	1
Centrarchidae	2	1	0
Plant material	1	0	0
Plant material + human-made structures	1	1	0
Cottidae	3	0	3
Plant material	2	0	2
Rock material + plant material	1	0	1
Cyprinidae	10	0	0
Waterbody creation	5	0	0
Plant material + waterbody modifications	4	0	0
Rock material + plant material	1	0	0
Esocidae	1	0	0
Waterbody creation	1	0	0
Gasterosteidae	1	0	0
Waterbody creation	1	0	0
Percidae	4	0	0
Waterbody creation	2	0	0
Plant material + waterbody modifications	2	0	0
Petromyzontidae	1	0	1
Plant material	1	0	1
Salmonidae	13	5	3
Rock material	5	4	1
Waterbody creation	3	0	0
Rock material + human-made structures	0	1	0
Rock material + plant material	2	0	2
Rock material + plant material + human-made structures	1	0	0
Rock material + plant material + waterbody modifications	2	0	0

longer-term evaluations, with 26% of the available data sets reporting restoration evaluations between 33 and 74 months after the last intervention was implemented, and 8% of data sets reported data more than 74 months after the last intervention was applied (Additional file 6).

Global meta-analyses

The overall mean weighted effect size for abundance was 0.54 (95% CI 0.32, 0.76; k = 39, p < 0.0001; Fig. 9), corresponding to a 60.4% (95% CI 47.51, 73.23) overall increase in substrate spawning fish abundance with spawning habitat creation or enhancement compared to controls. The majority of effect sizes were positive (i.e., *g* > 0; 31

(See figure on next page.)
Fig. 9 Summary plot of all effect size estimates from evaluations of the effectiveness of spawning habitat creation or enhancement on fish abundance (*k* = 39) subgrouped by intervention categories. Intervention type code definitions can be found in Table 3. Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in control areas (no intervention). Waterbody mods: waterbody modifications. Species Codes: BrkTrout: *Salvelinus fontinalis*; BrnTrout: *Salmo trutta*; Chinook: *Oncorhynchus tshawytscha*; Coho: *Oncorhynchus kisutch*; CmBream: *Abramis brama*; CmCarp: *Cyprinus carpio*; CmRoach: *Rutilus rutilus*; ERuffe: *Gymnocephalus cernuu*; EPerch: *Perca fluviatilis*; Gudgeon: *Gobio gobio*; LkTrout: *Salvelinus namaycush*; Lamprey: *Entosphenus tridentatus* and *Lampetra* spp.; LgBass: *Micropterus salmoides*; LnDace: *Rhinichthys cataractae*; Masu: *Oncorhynchus masou*; NPike: *Esox lucius*; PkSalmon: *Oncorhynchus gorbuscha*; RnTrout: *Oncorhynchus mykiss*; RtSculpin: *Cottus perplexus*; SlSculpin: *Cottus cognatus*; SmBass: *Micropterus dolomieu*; ShTrout: *Oncorhynchus mykiss*; Sunbleak: *Leucaspius delinineatus*; TsStick: *Gasterosteus aculeatus*; ToSculpin: *Cottus rhotheus*; various: various species; WhSturg: *Acipenser transmontanus*

Citation & Species Code	Intervention type	Hedges' g [95% CI]
Rock+Plant combinations		
Nagata 2002 Masu	LOG/GRAV/STRUC	0.33 [-2.37, 3.03]
Naito 2016d BrkTrout	LOG/ROCK	0.62 [-1.81, 3.05]
Naito 2016c LnDace	LOG/ROCK	-0.72 [-3.15, 1.72]
Naito 2016b RnTrout	LOG/ROCK	2.51 [-0.20, 5.21]
Naito 2016a SlSculpin	LOG/ROCK	-0.59 [-3.02, 1.84]
Ward 2003b Coho	NUTR/ROCK/LOG	0.33 [-2.08, 2.75]
Ward 2003a ShTrout	NUTR/ROCK/LOG	-0.67 [-3.10, 1.76]
Waterbody creation		
Pedersen 2009 GelsA_ BrnTrout	STREAM	0.33 [-1.28, 1.94]
Pedersen 2009 Stensbæk BrnTrout	STREAM	1.04 [-1.26, 3.34]
Pedersen 2009 RydsA BrnTrout	STREAM	0.92 [-1.36, 3.20]
Langler 2001o CmBream	BAY	0.70 [-0.52, 1.92]
Langler 2001n CmCarp	BAY	0.57 [-0.62, 1.76]
Langler 2001m CmRoach	BAY	-0.70 [-1.92, 0.52]
Langler 2001l ERuffe	BAY	-0.11 [-1.23, 1.01]
Langler 2001k EPerch	BAY	1.80 [0.12, 3.48]
Langler 2001j Gudgeon	BAY	-0.09 [-1.21, 1.03]
Langler 2001i NPike	BAY	-0.57 [-1.76, 0.62]
Langler 2001h Sunbleak	BAY	0.55 [-0.63, 1.73]
Langler 2001g TsStick	BAY	0.00 [-1.12, 1.12]
Toft 2013b various	BAY	2.82 [0.31, 5.34]
Moerke 2003 various	STREAM	1.05 [-0.29, 2.38]
Plant material		
Hunt 2002 LgBass	LOG	0.57 [-0.62, 1.76]
Roni 2003d Lamprey	LOG	0.50 [-0.16, 1.17]
Roni 2003c RtSculpin	LOG	0.57 [-0.01, 1.15]
Roni 2003b ToSculpin	LOG	0.18 [-0.50, 0.85]
Plant material+human-made structure		
Wills 2004 SmBass	LOG/STRUC	1.03 [-0.30, 2.35]
Plant material+waterbody mod		
Langler 2001f CmBream	RIP/VEG	0.57 [-0.62, 1.76]
Langler 2001e CmRoach	RIP/VEG	0.32 [-0.82, 1.46]
Langler 2001d ERuffe	RIP/VEG	0.09 [-1.03, 1.21]
Langler 2001c EPerch	RIP/VEG	0.69 [-0.53, 1.91]
Langler 2001b Gudgeon	RIP/VEG	0.54 [-0.64, 1.72]
Langler 2001a Sunbleak	RIP/VEG	0.57 [-0.62, 1.76]
Lorenz 2013 various	LOG/ EXCAV	0.44 [-0.02, 0.91]
Rock material		
Marsden 1995 LkTrout	ROCK	0.70 [-1.87, 3.28]
Roni 2008a Coho	ROCK	1.15 [0.15, 2.15]
Shackle 1999 BrnTrout	WASH/ WASH/ WASH	0.74 [-0.43, 1.91]
Zeug 2014b Chinook	GRAV	2.76 [0.95, 4.57]
Zeug 2014a ShTrout	GRAV	0.85 [-0.42, 2.13]
Toft 2013a various	SED	1.47 [-0.42, 3.35]
RE Model for abundance effect sizes (k=39)		0.54 [0.32, 0.76]

Hedges' g

of 39), with the remaining 20% showing neutral or negative responses (i.e., $g \leq 0$) to spawning habitat creation or enhancement; however, most of the individual effect sizes were not statistically significant, having large confidence intervals that overlapped zero (35 out of 39 effect sizes) (Fig. 9). The Q test for heterogeneity suggested that there was no statistically significant heterogeneity between effect sizes ($Q = 32.75$, p $= 0.711$). The funnel plot for the random effects model for abundance did not show an obvious pattern of publication bias; i.e., as study sample size increased, the variance in effect sizes decreased (see Additional file 7: Fig. S1). Also, the fail-safe number ($N = 237$) was greater than $5k + 10$ [($5 * 39 + 10) = 205$], suggesting the results from the random effects model was robust against potential publication bias (i.e., a relatively large number of studies was suggested to be required to eliminate the significant overall effect size).

The overall mean weighted effect size suggests an overall increase in substrate spawning fish survival with spawning habitat creation or enhancement compared to controls [Hedge's $g = 6.05$ (95% CI 0.13, 11.96), k $= 6$, p $= 0.045$; 26.41% (95% CI 0.84, 51.98); Fig. 10]. However, the sample size was quite small and three of the studies had very large positive effect sizes (lake trout [58], brown trout [59], and chinook salmon [10]), and as such may be having a disproportionately high impact on the mean effect size for improving survival (Fig. 10). The Q test for heterogeneity suggested that there was significant heterogeneity between effect sizes ($Q = 44.83$, p < 0.0001), suggesting that there is significant heterogeneity that could be explored using mixed effects meta-analysis models; however, given the sample size, the influence of

moderators could not be assessed due to the potential of overparameterization. The funnel plot for the random effects model for survival did not show an obvious pattern of publication bias; however, with this small number of studies, it is difficult to determine asymmetry (Additional file 7: Fig. S2). Furthermore, the failsafe number ($N = 0$) was not greater than 5k * 10 [($5 * 6 + 10) = 40$], suggesting the results from the random effects model may not be robust against potential publication bias.

The overall mean weighted effect size for body size was not statistically significant [Hedge's $g = 0.03$ (95% CI -0.29, 0.36), k $= 8$, p $= 0.84$; 0.48% (95% CI -3.27, 4.23); Fig. 11]. The result of the Q test also suggested that there was not significant heterogeneity in effect sizes between studies ($Q = 4.60$, p $= 0.709$). This was also supported by visual assessment of the forest plot for this meta-analysis, in which there were no individual studies with significant effect sizes (Fig. 11). The funnel plot was non-informative for this low number of studies and the failsafe number was 0 (Additional file 7: Fig. S3).

Effects of moderators on abundance

The following section addresses our secondary research questions (Fig. 12). There were too few effect sizes within the survival and body size subsets to permit meaningful analyses for these questions; therefore, all analyses below use a subset of fish abundance responses. For all analyses, we present the main results and plots in this section and summarize all outputs in Fig. 12 and Table 8.

1(a). To what extent does spawning habitat creation or enhancement effectiveness differ among inter-
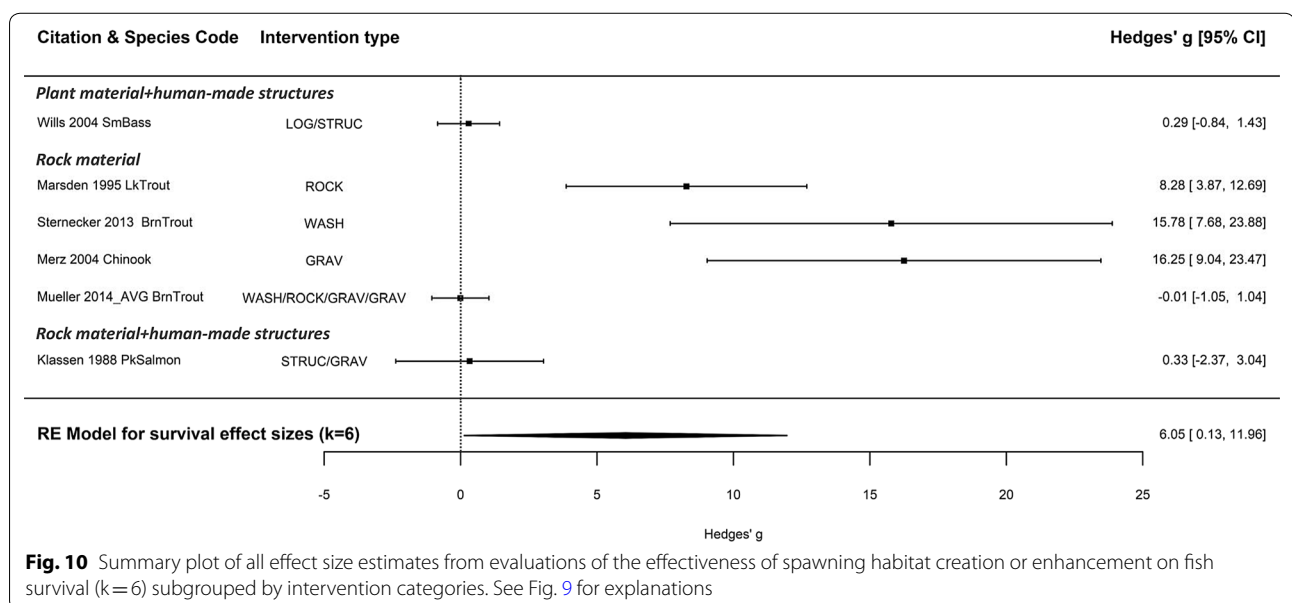
Fig. 10 Summary plot of all effect size estimates from evaluations of the effectiveness of spawning habitat creation or enhancement on fish survival (k $= 6$) subgrouped by intervention categories. See Fig. 9 for explanations
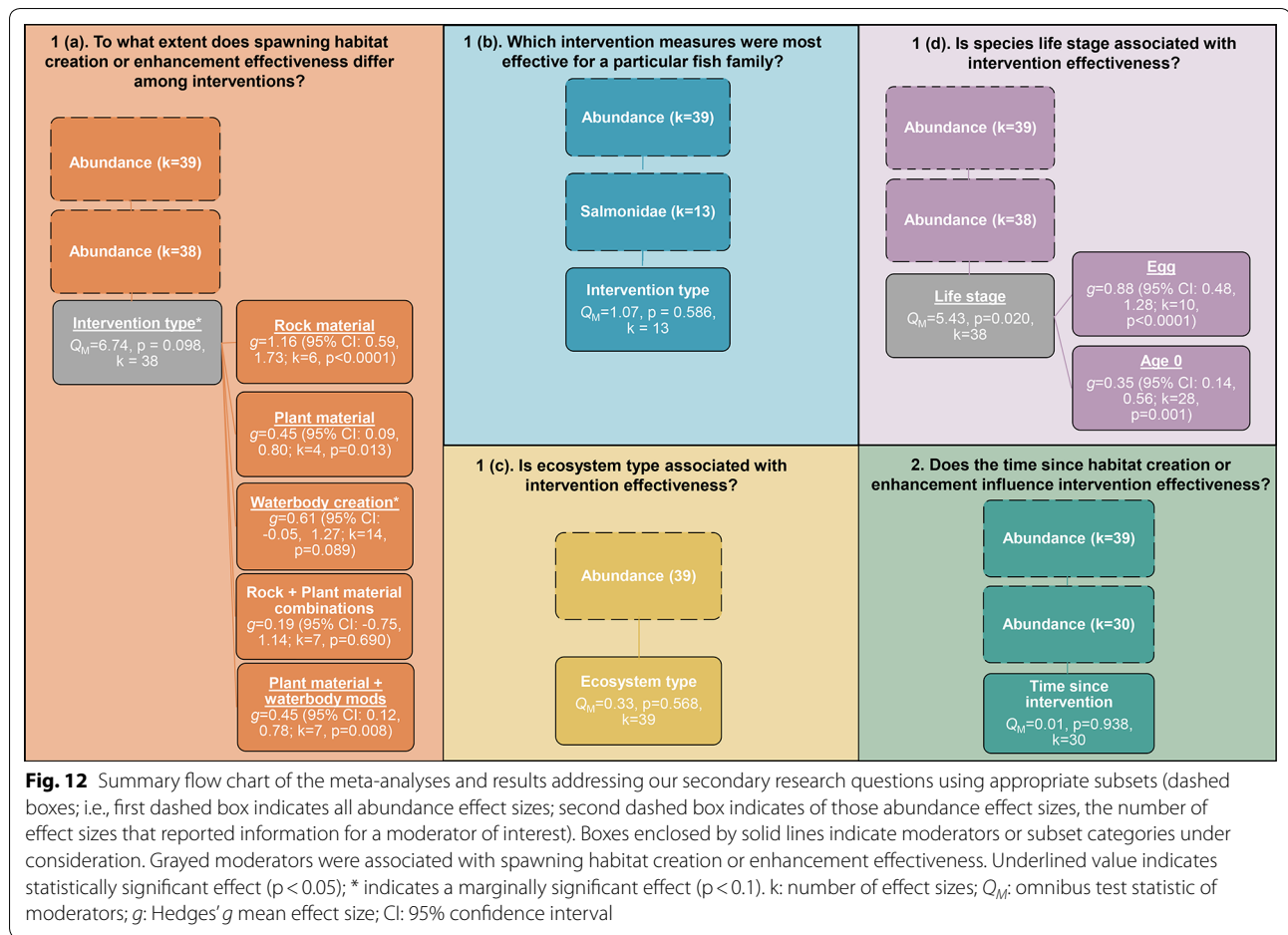
Fig. 11 Summary plot of all effect size estimates from evaluations of the effectiveness of spawning habitat creation or enhancement on fish body size ($k = 8$) subgrouped by intervention categories. See Fig. 9 for explanations
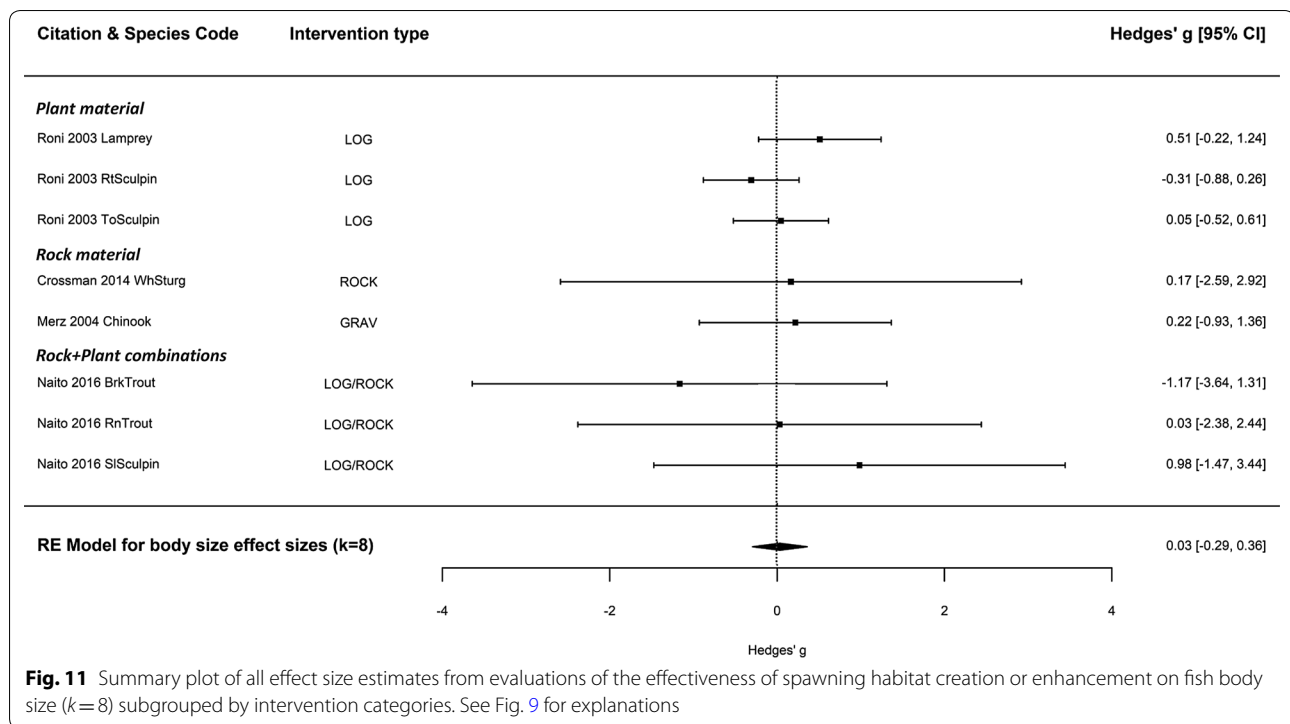
ventions?—There were only sufficient sample sizes (i.e., > 2 data sets from ≥ 2 independent studies) to include the following intervention types: (1) Rock material alone; (2) Plant material alone; (3) Waterbody creation alone; (4) Rock material + Plant material combinations; (5) Plant material + Waterbody modifications (Fig. 12). The effectiveness of spawning habitat creation or enhancement in increasing fish abundance varied among intervention types (Fig. 12, Table 8A; moderated model; and see Additional file 7: Fig. S4), though the influence was weak. Fish were more abundant with the addition or alteration of rock material, the addition of plant material (all log additions), and combining plant material with physical alterations to the waterbody (i.e., riparian modifications or excavation) than control sites, with a stronger positive effect for rock material (Figs. 12, 13).

1(b). Which intervention measures were most effective for particular fish families?—There was only sufficient sample size within the Salmonidae family to address this question, with the following intervention types: (1) the addition of rock material; (2) the creation of a new waterbody or extension of an exist-

ing waterbody, and (3) the combination of the addition of rock material + one or more different habitat creation/enhancement interventions. The effectiveness of spawning habitat creation or enhancement in increasing salmonid abundance did not vary among intervention types (Fig. 12 and Table 8B; and see Additional file 7: Fig. S5).

1(c). Is ecosystem type (lotic vs. lentic waterbodies) associated with intervention effectiveness?—We found no detectable effect of ecosystem type on average effect sizes (Fig. 12 and Table 8C; and see Additional file 7: Fig. S6).

1(d). Is species life stage associated with intervention effectiveness?—We detected a statistically significant effect of life stage on fish abundance (Table 8D), with the abundance of egg life stages associated with larger effect sizes than age-0 life stages (Figs. 12, 14; and see Additional file 7: Fig. S7).

2. Does the time since habitat creation or enhancement influence intervention effectiveness?—We found no detectable effect of time since intervention on average effect sizes (Fig. 12 and Table 8E; and see Additional file 7: Fig. S8).

Fig. 12 Summary flow chart of the meta-analyses and results addressing our secondary research questions using appropriate subsets (dashed boxes; i.e., first dashed box indicates all abundance effect sizes; second dashed box indicates of those abundance effect sizes, the number of effect sizes that reported information for a moderator of interest). Boxes enclosed by solid lines indicate moderators or subset categories under consideration. Grayed moderators were associated with spawning habitat creation or enhancement effectiveness. Underlined value indicates statistically significant effect (p < 0.05); * indicates a marginally significant effect (p < 0.1). k: number of effect sizes; Q_M: omnibus test statistic of moderators; g: Hedges' g mean effect size; CI: 95% confidence interval

Table 8 Summary results of meta-analyses using subsets of fish abundance effect sizes for testing the influence of the given moderator variable related to our secondary research questions

Moderator	k	Q statistic (p-value)	Q_M (p-value)	Q_E (p-value)
(A)				
Unmoderated model	38	32.10 (p = 0.698)	–	–
Intervention type	38	–	6.74 (p = 0.098)*	25.37 (p = 0.826)
(B)				
Unmoderated model	13	8.25 (p = 0.766)	–	–
Intervention type	13	–	1.07 (p = 0.586)	7.18 (p = 0.709)
(C)				
Unmoderated model	39	32.75 (p = 0.711)	–	–
Ecosystem type	39	–	0.33 (p = 0.568)	32.30 (p = 0.689)
(D)				
Unmoderated model	38	30.99 (p = 0.746)	–	–
Life stage	38	–	5.43 (p = 0.020)	25.56 (p = 0.902)
(E)				
Unmoderated model	30	24.55 (p = 0.701)	–	–
Time since intervention	30	–	0.003 (p = 0.957)	24.54 (p = 0.653)

Unmoderated model: random-effects model; k: number of effect sizes; Q statistic: value of homogeneity test; Q_M: omnibus test statistic of moderators; Q_E: unexplained heterogeneity; CI: 95% confidence intervals

Significance at p < 0.05; * Significance at p < 0.1
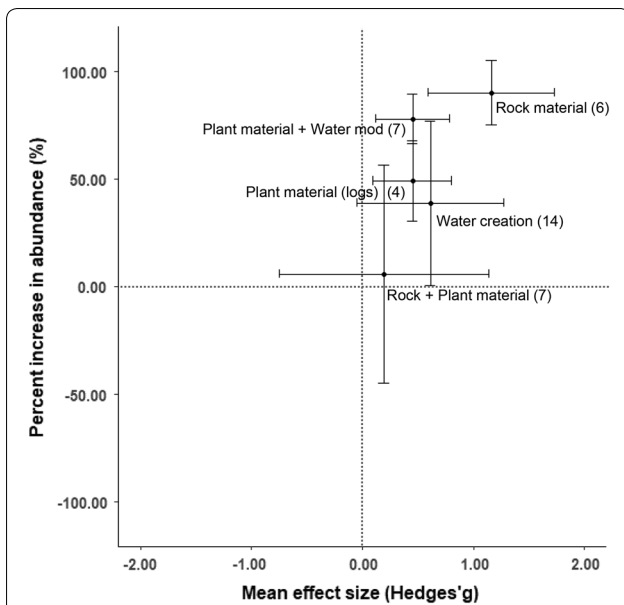
Fig. 13 Relationship between weighted-mean effect sizes and the weighted-mean percent increase in fish abundance for spawning habitat creation or enhancement interventions. Values in parentheses are the number of effect size estimates. Error bars indicate 95% confidence intervals. A positive mean value (right of/ above the dashed zero line) indicates that the abundance was higher in treatment areas than in control areas (no intervention). Water creation: creation of a new waterbody or extension of an existing waterbody; Water mod: waterbody modification
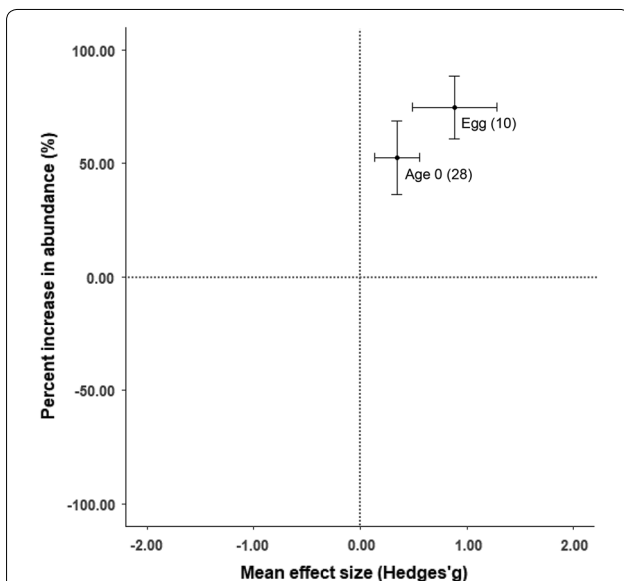
Fig. 14 Relationship between weighted-mean effect sizes and the weighted-mean percent increase in fish abundance for egg and age-0 life stages. Values in parentheses are the number of effect size estimates. Error bars indicate 95% confidence intervals. A positive mean value (right of/above the dashed zero line) indicates that the abundance was higher in treatment areas than in control areas (no intervention)

Discussion

Although the effectiveness of restoration or alteration of aquatic habitat has previously been reviewed (e.g., [17, 28–32]), our systematic review greatly improves on past reviews by providing the most extensive, systematic search on the effectiveness of habitat creation or enhancement for substrate-spawning fish. The previous lack of comprehensive reviews on this topic is likely due in part to the nature of fish habitat restoration (e.g., often conducted by grassroots organizations and volunteers) and the fact that many projects lack proper monitoring. However, many other restoration projects are undertaken to fulfil regulatory requirements where monitoring could be mandatory [60–62]. Moreover, data are often not published or are difficult to find [3]. For this review, we systematically obtained all available literature on habitat creation or enhancement for substrate-spawning fish, and as a result, have an extensive database that contains studies for several species, habitat types, restoration types, and locations around the world.

We identified 75 relevant studies, of which only 22 were eligible for quantitative analysis (all medium-validity studies). We acknowledge that our review does not represent the whole knowledge base on the subject. For instance, we excluded many unreplicated studies (i.e., only one treatment and/or one control site) that were ineligible for quantitative analysis, but did contribute to the narrative review. Furthermore, during our screening process, several articles (28; see Fig. 1) were excluded from this review completely due to lack of proper comparator (i.e., before or control data). These studies described above could contribute useful information on this topic; however, the use of the systematic review approach to evaluate the existing literature base allowed us to identify to most relevant, and reliable (minimum biased) studies using this rigorous, objective, and transparent methodology.

Effectiveness of interventions

Overall, spawning habitat creation or enhancement generally resulted in higher values of some biological metrics (fish abundance and survival) than control areas not receiving any habitat creation or enhancement. Although many of the effects within the individual studies were not statistically significant (i.e., having 95% confidence intervals that overlapped zero), a meaningful pooled effect (e.g., meta-analytically pooled effect) can arise [63] by examining the overlap of confidence intervals of the effect sizes across the individual studies. Here again, the purpose of these summary effect sizes was to identify general trends in the evidence base. Although we found little between-estimate heterogeneity in mean fish abundance in response to habitat creation/

enhancements—suggesting a fairly consistent response to interventions—interpretation of summary effects should include some consideration for context. In particular, our analyses were limited by a small number of effect size estimates in general, and by a taxonomic bias towards salmonids (i.e., salmonids were 1 of 9 families represented in the abundance meta-analysis but were ~40% of data sets). As such, we caution interpretation of summary effects as evidence that any habitat creation or enhancement measure designed for substrate spawning fish will improve a fish response such as abundance. That said, our findings are consistent with several other reviews that showed a positive effect of habitat enhancements or alterations on fish at various life stages [7, 9, 17, 64].

Furthermore, we discourage the use of our alternative effect size metric (i.e., weighted-mean percent change in intervention effectiveness) beyond its intended use as interpretational aid to accompany the Hedges' *g* measure. While this metric was found to be closely related to effect size estimates for some fish outcomes in this review (i.e., abundance: $r = 0.748$, $p < 0.0001$, $k = 39$; survival: $r = 0.625$, $p = 0.184$, $k = 6$; body size: $r = 0.660$, $p = 0.075$, $k = 8$), it is unclear, in a broad sense, how closely related this metric is to Hedges' *g*. Data within intervention types were often negatively skewed and these distributions were not improved by applying a transformation. Uncertainty estimation using bootstrapping was not appropriate given the small sample sizes, which has the potential to increase uncertainty in estimated weighted-mean percent changes and confidence intervals [65, 66]. As such, it would not be appropriate to explicitly use the estimated weighted-mean percent change in intervention effectiveness towards advocating an offset ratio. For example, a comparison between an estimated 90% increase in abundance with the addition of rock material relative to areas with no intervention and a 49% increase in abundance with the addition of logs relative to control sites, should not be used to infer that fewer rocks than logs are required to achieve equivalency. Furthermore, the weighted-mean percent change in intervention effectiveness should not be used as a benchmark indicator for effectiveness consideration (e.g., a 90% increase in abundance must be achieved for rocks to be considered effective and anything less would require additional offsetting). For both metrics (i.e., Hedges' *g* and percent change in intervention effectiveness), all we can infer is the direction (an increase, decrease, or no change) and the relative strength of the treatment effect; neither metric can explicitly provide an offsetting value (e.g., 100 m^2 of rock are needed to achieve an increase in age-0 fish density). As such, we base our conclusions on the relative effectiveness of spawning habitat creation or enhancement interventions for substrate-spawning fish using

the Hedges' *g* effect size estimates, and strictly use the weighted-mean percent change in intervention effectiveness as a coarse, supplemental indicator of the relative magnitude of the treatment effect. Therefore, we can infer from our results that the addition of rocks is more likely to provide a greater benefit to substrate spawning fishes than an offset based on logs, and while the actual size of the offset should be dependent on the impact and ecological context, this knowledge reduces uncertainty in the ultimate outcome, providing greater confidence in its application.

The effectiveness of spawning habitat creation or enhancement in increasing fish abundance varied among intervention types. The addition or alteration of rock material was effective in increasing the abundance of substrate-spawning fish compared to controls (Figs. 12, 13). There was a strong taxonomic bias towards salmonids for this intervention, i.e., 5/6 data sets for abundance were salmonids [Hedges' *g* for salmonids alone = 1.13 (95% CI 0.53, 1.73); k = 5; 89.79% percent change in intervention effectiveness (95% CI 72.87, 106.72) vs. Hedges' *g* for all rock data sets = 1.16 (95% CI 0.59, 1.73); k = 6; 90.23% (95% CI 75.02, 105.43)], and studies only focused on egg or larvae life stages (i.e., no studies focused on adult fish) (Table 6). This result is not surprising given the presumed benefits of new or clean gravel for the creation of redds for salmonids (i.e., easier excavation, less fine sediment; [67]). However, the data set limits our ability to draw conclusions on other substrate-spawning fish. Furthermore, we could not quantitatively investigate the relative effectiveness of different forms of rock material (e.g., cobble vs. gravel sized rock material) due to small sample sizes within these finer scale categories.

The addition of plant material (e.g., large woody debris, planting of macrophytes) with or without physical alterations to the waterbody (i.e., riparian modifications or excavation) was also effective in increasing substrate-spawning fish abundance on average compared to controls (Figs. 12, 13). Studies evaluating the effectiveness of adding plant material have targeted centrarchids (i.e., *Micropterus dolomieu* and *M. salmoides*; smallmouth and largemouth bass), cottids (i.e., *Cottus perplexus* and *C. rhotheus*; reticulate and torrent sculpins), cyprinids (i.e., *Abramis brama*, *Gobio gobio*, *Leucaspius delinineatus*, and *Rutilus rutilus*; common bream, gudgeon, sunbleak, common roach), percids (i.e., *Gymnocephalus cernua* and *Perca fluviatilis*; Eurasian ruffe and European perch), and petromyzontids (i.e., *Entosphenus tridentatus* and *Lampetra* spp.; lamprey) species; however, they have not targeted salmonids. Plant material (i.e., wood) has been well studied in the context of salmonids but in terms of providing rearing/growth/refuge habitat for juveniles and adults (see [32, 68]) rather than its direct role in

enhancing substrate for spawning given that salmonids exclusively reproduce on rock substrates.

There was little detectable evidence that other measures, including the creation of a new waterbody or the extension of an existing waterbody, or combining additions of rock and plant material, along with other interventions, increased fish abundance (Figs. 12, 13). Studies evaluating the effectiveness of combining additions of rock and plant material targeted salmonids, cottids, and cyprinids and presented data from only age-0 fish, which may explain the differences in the magnitude of intervention effectiveness observed for these interventions when implemented alone versus in combination (i.e., targeting different species/life stages). Interestingly, this observation does not appear to be due to differences in the types of plant material used when applied alone or in combination, as logs were added in all cases. However, the intent behind different forms of restoration (e.g., rock and plant material) was not always reported in papers so it may be the case that different materials were used for different purposes (e.g., plant material was used to stabilize a shoreline and reduce erosion while rock was added to provide actual spawning habitat).

In situations where different intervention types are used in combination to enhance or create habitat, it was not possible to analyze the relative effectiveness of the individual intervention types (e.g., effect of only rock materials when combined with plant materials). In the case of artificial streams, for example, there were often several interventions performed simultaneously to create the new waterbody (i.e., excavation, addition of gravel, planting of macrophytes). Though we recognize that all components may play significant roles in the success or failure of the restoration, in such situations we were unable to isolate the effects of the individual interventions within the study and had to therefore treat them as a single intervention category.

Biotic responses to particular restoration techniques are highly context-dependent. For instance, physical spawning habitat may be created for salmonids by depositing rocky material, but the quality of such habitat depends on many other factors aside from the composition of the material itself. These include thermal conditions, flow, sedimentation, and dissolved oxygen levels [69–71]. Fish may not use created habitats unless all attributes are suitable, and only partly suitable artificial habitats could inadvertently serve as ecological sinks, especially if they fill with sediment after redds are constructed. Unless hatching rates are monitored, the full effectiveness of spawning habitats may be obscured. Even when created habitats are highly suitable, increases in abundance will only occur if fish can access them [72]. Barriers located beyond the restoration site could prevent use and therefore any increase in abundance. Finally, the magnitude of increase in abundance depends on what currently limits population size. If spawning habitat is a limiting factor, then habitat creation projects may be highly successful. However, other exogenous factors such as current population size, prey abundance, and climate can regulate biotic responses to even perfectly constructed and located habitats. In depleted populations for instance, recruitment is limited by the number of available spawners, and the rate of increase in abundance depends on the initial population size [73]. Habitat restoration or enhancement projects should involve careful consideration of the local context at the planning stage to maximize probability of achieving objectives.

We were severely limited in our ability to draw conclusions on the effectiveness of habitat creation or enhancement measures for particular fish taxa since there have been relatively few studies within and across different restoration measures. For salmonids, the effectiveness of spawning habitat creation or enhancement in increasing salmonid abundance did not differ among intervention types (Fig. 12 and Table 8B).

Reasons for heterogeneity

As mentioned above, for fish outcomes, there was little variation between effect size estimates, indicating a certain degree of consistency in fish responses to habitat creation or enhancement measures. The reason for this limited heterogeneity in the observed estimates of habitat creation or enhancement effectiveness is somewhat unclear, given the variety of interventions, ecosystem types, and response metrics used. Though we restricted our review to include studies conducted in temperate regions only, there was considerable variation in studied environments. For example, although half of our abundance effect sizes were from studies conducted in North America, these studies spanned a variety of climate zones within both the United States (e.g., Köppen-Geiger climate zones as defined by Peel et al. [74]: Cfa, Csb, Dfa) and Canada (e.g., Dfb, Dfc, Cfb). Therefore, this low level of heterogeneity does not appear to be due to a lack of regional variation among study systems.

Low heterogeneity across effect size estimates may be explained, at least in part, by a research focus on a relatively small number of fish species. For example, in the quantitative analyses, there were only abundance data for 24 species, from 17 genera and 8 families, of which, a third of all species were from a single family (i.e., Salmonidae). This observation not only highlights a clear taxonomic bias in the current literature base but also a potential consistency in species responses to habitat creation or enhancement projects.

Knowledge gaps and clusters

Overall, there were few studies included in the quantitative synthesis that investigated habitat creation or enhancement interventions either alone, or in combination. For example, when combining across all fish taxa for the quantitative synthesis, the greatest number of data sets for any intervention type was 14, involving varying taxa and ecosystem types. This essentially precluded us from drawing any strong conclusions about the effectiveness of measures for fish habitat restoration.

Small sample size limited our ability to investigate the influence of certain variables that could affect the success of different habitat creation or enhancement measures in a robust manner (e.g., time since the intervention was applied or species-specific factors [75]). There was often little variation among estimates in the extent to which techniques increased fish abundance within models, suggesting the effectiveness of these techniques was relatively consistent across the studies included in this review (i.e., given the species and interventions for which there were sufficient data).

The majority of the research we examined focused on a small number of fish species and families. Of all studies included the narrative synthesis, 63.7% reported on the effect of restorations on salmonids and a taxonomic bias towards salmonids can been seen throughout this review. This is perhaps not surprising given the cultural, economic, and recreational significance of salmonids [76, 77] and the resources dedicated to their conservation. Further research is needed on a broader range of substrate-spawning fish with a particular focus on those that are at risk or considered a target of restoration activities, both of which can vary significantly by jurisdiction.

Limitations of the review and evidence base

Collectively, the studies reviewed here did not provide insight on population-level responses to spawning habitat restoration. For instance, it was unclear whether the amount of existing spawning habitat was a limiting factor, and whether created habitats simply attracted and relocated fish, or increased the overall productive capacity of the ecosystem. The attraction-production debate for habitat restoration has been discussed for decades (see [78–80]). More recently in the context of salmonid habitat restoration, Roni [81] reviewed the complex relationships between restoration and fish movement, abundance, and survival and recommended more detailed monitoring at an ecosystem-scale. This research at both the reach and watershed scale is required to identify limiting factors, assess population changes, and differentiate attraction from additionality.

Furthermore, our review focuses only on early life stages (narrative and quantitative syntheses) and adult spawners (narrative synthesis only), which can make it difficult to assess population-level effects of restoration efforts. We excluded outcomes related to juveniles to avoid confounding the effects of spawning habitat enhancements with those of nursery or rearing habitat quality, where factors such as temperature, nutrient availability, food, and cover can dictate success (e.g., the critical-period concept; [82, 83]). This decision was made because it becomes increasingly difficult to assess the effects of spawning habitat restoration on success at later life stages, whereas embryonic survival can often be taken as direct evidence of successful restoration [28]. A certain level of ambiguity exists however around cases where authors report on age-0 fishes for evaluations of spawning habitat creation or enhancements. For the purpose of this review, we included all age-0 fish responses based on the assumption that if authors used this age to measure the response, it was likely the most relevant, practical, and/or appropriate age for that species/study. While there are instances when this assumption is likely valid (i.e., where it is known that this age class remains on the spawning substrate for a time period prior to relocation to nursery habitat, or sampling dispersing fish immediately on or downstream of spawning habitat creation/enhancements), there could also be cases where it is not and the age metric is more indicative of nursery habitat use in that the fish could have hatched elsewhere. Although it would have been informative to determine the influence of including all age-0 fish outcomes on the effectiveness of interventions, the evidence base was not large enough to allow us to undertake such a sensitivity analysis (i.e., compare summary effect size with and without the inclusion of age-0 fish responses).

Due to limitations in the data, we were unable to analyze the long-term effect of habitat restoration or enhancement on substrate-spawning fish. The majority of studies in this review were based on short-term monitoring. Long-term studies are important to identify changes in the effectiveness and longevity of the interventions. For example, gravel beds may wash downstream or fill with silt, only resulting in a positive effect for a year or two. Conversely new or enhanced habitat may increase in value over time as it is naturalized or if it takes time for fish to find and use it. Restoration actions are often coupled with short-term monitoring rather than being the focus of long-term experiments, and resources are often prioritized towards action rather than scientific assessment of effectiveness [84, 85]. Of all studies included in the narrative synthesis, most reported only one to two years of post-monitoring (Fig. 7) and the only studies reporting greater than eight years stemmed from two articles [54, 55]. This trend was also observed in a recent meta-analysis on the effect of instream structures on

salmonids that found fewer than five projects were monitored beyond 10 years [32]. To properly assess the effectiveness of offsetting activities, it is recommended that a long-term BACI design with a minimum of three years of before data and blocks of *continuous* after monitoring be used (e.g., three continuous years of sampling immediately post-treatment, and additional three years of sampling at a later time (e.g., four to six years post-treatment), and revisit 10 years post-treatment [86, 87]).

Though this review did aim to include temperate marine and freshwater environments, there were no marine studies that met our inclusion criteria. Spawning habitat restoration occurs in coastal environments; however, such activities are perhaps less frequently conducted or monitored using an experimental design that met our strict criteria. It is also more common for coastal restoration to focus on improving nearshore habitats for juveniles and thus the outcome metric reported would not have been captured by our search strings as relevant to spawning activity.

Our review was limited to only English articles. Though there may be valuable articles, particularly grey literature, from other countries that are not published in English, we feel that we have captured what is available and most relevant given the Canadian (or more broadly, North American) context of this review.

There was limited evidence of publication bias; however, there were some geographical and taxonomic biases in the data included in quantitative synthesis. The majority of studies included in quantitative synthesis were from North America (56.6%) and a large percentage (39.6%) targeted salmonids (78.1% and 63.7%, respectively for studies included in the narrative synthesis). Though we did search for available grey literature through websites, the Advisory Team, evidence call-outs, and social media, few relevant articles were obtained and it is almost certain that additional grey literature exists. Habitat restoration is sometimes performed by groups that do not have the resources to publish their results or perform long-term monitoring (e.g., practitioners focused on implementing restoration). It is possible that many habitat restoration effectiveness monitoring activities go undocumented, or are reported in internal documents that were not accessible to our review team.

As mentioned previously, several articles were excluded from this review due to lack of a proper comparator. Though these excluded studies provide insight into habitat restoration practices (e.g., what interventions are used for specific species or regions), their assumed successes and/or failures cannot be used to determine effectiveness and therefore cannot contribute to the context of this review. Additionally, 29 studies (from 24 articles) were included in the narrative component of this review, as

they were unreplicated and therefore ineligible for quantitative analysis. Poor study design has previously been noted specifically in a Fisheries and Oceans Canada context by Harper and Quigley [88] who examined habitat compensation authorizations and found that only 56% of projects had pre-treatment assessment methods that matched those of post-treatment, making it difficult to track the effectiveness of projects being completed. Several assessment methods exist for monitoring the success of spawning habitat restoration. Although some may be labour intensive [28], others require little technical expertise and could be implemented to better assess the effectiveness of projects. As a consequence, we recommend that researchers and practitioners ensure that they collect data either from before the intervention, or from a reference area nearby to measure the effect of no intervention and include replication at the level of the intervention whenever possible.

However, given that many real-world restorations or offsets will be single interventions at one site, at a minimum it is important to sample multiple locations at that intervention to provide a variance allowing for quantitative analysis. This pseudoreplication would need to be acknowledged and accounted for in meta-analyses to ensure such data are not over-weighted relative to a true independently replicated intervention. In our study, 20 of the 53 data sets included in our quantitative analyses were based on partly subsampled or pseudoreplicated data. In these instances, outcome means and variances were not from independent replicates but subsamples such as subplots or at the nest level. Although we made a quantitative adjustment to avoid giving pseudoreplicated data too much weight in analysis (see Additional file 5; but also see [89, 90]), we were unable to test the impact of these data (i.e., through sensitivity analysis) on our findings because of small sample size. It remains important that robust monitoring be conducted at the single intervention level, and for such studies to be published acknowledging the pseudoreplication. In combination, unreplicated studies can ultimately contribute to the knowledge base and improve the sample size for a pooled analysis or adjusted meta-analysis, particularly if interventions are similar among studies, and monitoring protocols are consistent.

Moreover, several studies included in this review had poor data reporting limiting our ability to use them in quantitative analysis. For example, 17 studies did not report variance of group means or provide the raw data necessary for such calculations. There were also six studies that averaged data across years without providing individual year data which impeded our ability to isolate the last sampling year as outlined in our data extraction strategy. Other common reporting issues included

unclear study timelines (i.e., we were unable to accurately determine time since intervention), and unclear sampling and analysis units [i.e., sample sizes were not always reported or they could not be determined from reported statistical results (or lack thereof)]. To better facilitate quantitative syntheses, we recommend that authors provide raw data either directly in the article or an appendix/data archiving site, for each year, species, intervention type, and control and impact site separately. In other words, data should not be combined across years and/or sites and authors should clearly distinguish before, during, and after intervention time periods when applicable.

Implications for policy and research

Availability of adequate spawning habitat is critical for the sustainability of some fish populations. Attempts to restore or create this habitat requires planning to ensure resources are being used appropriately and the goals of the project are achieved. To better inform regulators and habitat practitioners, there is a dire need for improvement in research for a broader range of habitat creation and enhancement measures (e.g., artificial streams or bays, human-made structures, waterbody modifications). We recommend, as others have before [85, 88, 91, 92], some specific opportunities for improving the evidence base (both in terms of quantity and quality) on this topic that include:

- Consider requiring that all aquatic habitat enhancement activities include a monitoring component, which ideally includes a replicated intervention, before and after comparison, and continues for at least three years, particularly when activities are undertaken to fulfill regulatory requirements. Monitoring should be encouraged, but perhaps not required, for voluntary enhancement activities.
- Develop training programs that build capacity for conducting more effective monitoring within the communities that engage in restoration activities (e.g., community groups, stewardship organizations, NGOs, practitioners within government).
- Create standardized databases or other means of collecting, aggregating and archiving monitoring data of effectiveness emanating from restoration projects.
- Develop monitoring standards (and funding mechanisms to support them) that enable potential inclusion in future systematic reviews - that is, monitoring requires sufficient rigour to pass critical appraisal (e.g., replicated and controlled experiments).
- Encourage practitioners with long-term data to analyze their results and share more widely—there is an increasing number of journals, such as Conservation Evidence, that include "case reports" for practitioners

to report on their work in a concise and narrative format that is attainable for non-researchers.
- Consider developing "big science" projects across different landscape types (and/or fish communities, ecosystems, and so on) that enable a comparative approach to restoration effectiveness using a standardized (and robust) monitoring/science framework. Active adaptive management is an ideal tool for assessing restoration effectiveness. This involves learning from the management of ecosystems by undertaking management actions as deliberate and ideally replicated experiments that test predicted outcomes [93].

General conclusions

In this review, we investigated the effect of spawning habitat creation or enhancement on substrate-spawning fish. This is of particular importance because habitat creation or enhancement is commonly used as an offsetting technique intended to increase fisheries productivity and counterbalance the effects of human development or activities that cannot be avoided or mitigated. The synthesis of available evidence suggests that the addition or alteration of rock material (e.g., addition of gravel, rocks, and boulders, substrate washing) was a consistently effective means of enhancing spawning habitat, but results may only be applicable for salmonids. Furthermore, synthesis suggests that on average, the addition of plant material with or without waterbody modifications was also effective at increasing fish abundance.

Overall, we were limited in our ability to address many of the questions that stakeholders had about the effectiveness of habitat creation or enhancement, in particular, questions related to species specific responses or the relative effectiveness of finer scale details of intervention types (i.e., is the addition of cobble more effective in increasing fish abundance than the addition of gravel?). We believe this is because of two main issues with the current literature base: (1) low study validity, and (2) limited replication of studies across species and interventions. Before we can provide recommendations with a higher level of certainty on the effectiveness of habitat creation or enhancement on substrate-spawning fish, we need to improve research and reporting, and expand our focus to include a broader range of species and intervention types. We provide several recommendations aimed at researchers and practitioners and recognize that they are most relevant to jurisdictions with an appropriate governance framework and the scientific, management and regulatory capacity to do so.

Taylor *et al. Environ Evid*   (2019) 8:19

Page 29 of 31

Additional files

Additional file 1. Database literature search strategies and results.

Additional file 2. List of articles excluded on the basis of full-text assessment and reasons for exclusion.

Additional file 3. Data-extraction sheet.

Additional file 4. List of supplementary articles.

Additional file 5. Data extraction considerations and additional calculations for meta-analysis.

Additional file 6. Meta-analysis database.

Additional file 7. Quantitative synthesis and publication bias.

Additional file 8. ROSES systematic review checklist.

Authors' contributions
This review is based on a draft written by JJT and TR. JJT performed searches, screened identified records and extracted data. TR performed all quantitative analyses. All authors assisted in editing and revising the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Results of literature searches, a list of articles excluded at full-text and supplementary articles, as well as the systematic review database, raw data, and analysis details are included as additional files with this report. A ROSES form [94] for this systematic review report is included as Additional file 8.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Canadian Centre for Evidence-Based Conservation, Institute of Environmental Sciences and Interdisciplinary Sciences, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada. [2] Fish Ecology and Conservation Physiology Laboratory, Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada. [3] Department of Biology and Institute of Environmental Science and Interdisciplinary Sciences, Carleton University, 1125 Colonel by Drive, Ottawa, ON, Canada. [4] Great Lakes Laboratory for Fisheries and Aquatic Science, Fisheries and Oceans Canada, 1219 Queen Street East, Sault Ste. Marie, ON, Canada. [5] Canadian Wildlife Federation, 350 Michael Cowpland Drive, Kanata, ON, Canada. [6] Central and Arctic Region, Fisheries and Oceans Canada, 501 University Cres, Winnipeg, MB, Canada. [7] Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada, 80 East White Hills, St. John's, NF, Canada. [8] Department of Biological Sciences, University of Alberta, 11455 Saskatchewan Drive, Edmonton, AB, Canada. [9] Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada.

References
1. Lapointe NWR, Cooke SJ, Imhof JG, Boisclair D, Casselman JM, Curry RA, Langer OE, McLaughlin RL, Minns CK, Post JR, Power M, Rasmussen JB, Reynolds JD, Richardson JS, Tonn WM. Principles for ensuring healthy and productive freshwater ecosystems that support sustainable fisheries. Environ Rev. 2014;22:1–25.
2. Rosenfeld JS, Hatfield T. Information needs for assessing critical habitat of freshwater fish. Can J Fish Aquat Sci. 2006;63(3):683–98.
3. Loughlin KG, Clarke KD. A review of methods used to offset residual impacts of development projects on fisheries productivity. Department of Fisheries and Oceans, Canadian Science Advisory Secretariat Research Document; 2014. pp. 97 vi+72p.
4. Balon EK. Reproductive guilds of fishes: a proposal and definition. J Fish Board Can. 1975;32(6):821–64.
5. Balon EK. Additions and amendments to the classification of reproductive styles in fishes. Environ Biol Fish. 1981;6(3–4):377–89.
6. Sass GG, Kitchell JF, Carpenter SR, Hrabik TR, Marburg AE, Turner MG. Fish community and food web responses to a whole-lake removal of coarse woody habitat. Fisheries. 2006;31(7):321–30.
7. Roni P, Beechie T, Pess G, Hanson K. Wood placement in river restoration: fact, fiction, and future direction. Can J Fish Aquat Sci. 2015;72(3):466–78.
8. Barlaup BT, Gabrielsen SE, Skoglund H, Wiers T. Addition of spawning gravel—a means to restore spawning habitat of atlantic salmon (*Salmo salar* L.), and Anadromous and resident brown trout (*Salmo trutta* L.) in regulated rivers. River Res Appl. 2008;24(5):543–50.
9. Keeley ER, Slaney PA. Quantitative measures of rearing and spawning habitat characteristics for stream-dwelling salmonids: guidelines for habitat restoration. Watershed Restoration Program, Ministry of Environment, Lands and Parks. 1996.
10. Merz JE, Setka JD, Pasternack GB, Wheaton JM. Predicting benefits of spawning-habitat rehabilitation to salmonid (*Oncorhynchus* spp.) fry production in a regulated California river. Can J Fish Aquat Sci. 2004;61(8):1433–46.
11. Crispin V, House R, Roberts D. Changes in instream habitat, large woody debris, and salmon habitat after the restructuring of a coastal Oregon stream. N Am J Fish Manag. 1993;13(1):96–102.
12. House R. An evaluation of stream restoration structures in a coastal Oregon stream, 1981–1993. N Am J Fish Manag. 1996;16(2):272–81.
13. Geist DR, Hanrahan TP, Arntzen EV, McMichael GA, Murray CJ, Chien YJ. Physicochemical characteristics of the hyporheic zone affect redd site selection by chum salmon and fall Chinook salmon in the Columbia River. N Am J Fish Manag. 2002;22(4):1077–85.
14. Snickars M, Sundblad G, Sandström A, Ljunggren L, Bergström U, Johansson G, Mattila J. Habitat selectivity of substrate-spawning fish: modelling requirements for the Eurasian perch *Perca fluviatilis*. Mar Ecol Prog Ser. 2010;398:235–43.
15. Miller B, Kendall A. Fish reproduction. In: Miller B, Kendall A, editors. Early life history of marine fishes. Berkeley: University of California Press; 2009. p. 9–38.
16. Curry RA, Noakes DL. Groundwater and the selection of spawning sites by brook trout (*Salvelinus fontinalis*). Can J Fish Aquat Sci. 1995;52(8):1733–40.
17. Smokorowski KE, Pratt TC. Effect of a change in physical structure and cover on fish and fish habitat in freshwater ecosystems—a review and meta-analysis. Environ Rev. 2007;15:15–41.
18. Fisheries and Oceans Canada. Assessment of measures to assess compensation and mitigation as related to the creation, rehabilitation, or restoration of spawning habitat for fluvial or lacustrine spawning salmonines. DFO Can Sci Advis Rep 2013/097. 2014.
19. Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Lévêque C, et al. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev. 2006;81(02):163–82.
20. Vörösmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, et al. Global threats to human water security and river biodiversity. Nature. 2010;467(7315):555–61.

Taylor *et al. Environ Evid*   (2019) 8:19

Page 30 of 31

21. Pullin AS, Stewart GB. Guidelines for systematic review in conservation and environmental management. Conserv Biol. 2006;20(6):1647–56.
22. Collaboration for Environmental Evidence. In: AS Pullin, GK Frampton, B Livoreil & G Petrokofsky, editors. Guidelines and standards for evidence synthesis in environmental management. Version 5.0. 2018. http://www.environmentalevidence.org/information-for-authors. Accessed 31 May 2018.
23. Fisheries and Oceans Canada. Fisheries productivity investment policy: a proponent's guide to offsetting. 2013. http://www.dfo-mpo.gc.ca/pnw-ppe/offsetting-guide-compensation/offsetting-guide-compensation-eng.pdf. Accessed 2 June 2018.
24. Fisheries and Oceans Canada. Science advice on offsetting techniques for managing the productivity of freshwater fisheries. DFO Can Sci Advis Rep 2013/074. 2014.
25. Rice J, Bradford MJ, Clarke KD, Koops MA, Randall RG, Wysocki R. The science framework for implementing the fisheries protection provisions of Canada's fisheries Act. Fisheries. 2016;40(6):268–75.
26. Haddaway NR, Kohl C, da Silva NR, Schiemann J, Spök A, Stewart R, Sweet JB, Wilhelm R. A framework for stakeholder engagement during systematic reviews and maps in environmental management. Environ Evid. 2017;6(1):11.
27. Macura B, Byström P, Airoldi L, Eriksson BK, Rudstam L, Støttrup JG. Impact of structural habitat modifications in coastal temperate systems on fish recruitment: a systematic review. Environ Evid. 2019;8(1):14.
28. Fitzsimons JD. Assessment of measures to assess compensation and mitigation as related to the creation, rehabilitation, or restoration of spawning habitat for fluvial or lacustrine spawning salmonines. Canadian Science Advisory Secretariat. 2014.
29. Roni P, Hanson K, Beechie T. Global review of the physical and biological effectiveness of stream habitat rehabilitation techniques. N Am J Fish Manag. 2008;28(3):856–90.
30. McLean M, Roseman EF, Pritt JJ, Kennedy G, Manny BA. Artificial reefs and reef restoration in the Laurentian Great Lakes. J Great Lakes Res. 2015;41(1):1–8.
31. Grossman GD, Jones GP, Seaman WJ Jr. Do artificial reefs increase regional fish production? A review of existing data. Fisheries. 1997;22(4):17–23.
32. Whiteway SL, Biron PM, Zimmermann A, Venter O, Grant JW. Do in-stream restoration structures enhance salmonid abundance? A meta-analysis. Can J Fish Aquat Sci. 2010;67(5):831–41.
33. Taylor JJ, Rytwinski T, Bennett JR, Smokorowski KE, Cooke SJ. The effectiveness of spawning habitat creation or enhancement for substrate spawning temperate fish: a systematic review protocol. Environ Evid. 2017;6(1):5.
34. Gebhards SY. Emergence and mortality of chinook salmon fry in a natural redd. Prog Fish Cult. 1961;23:91.
35. Porter TR. Fry emergence trap and holding box. Prog Fish Cult. 1973;35:104–6.
36. Field-Dodgson MS. Emergent fry trap for salmon. Prog Fish Cult. 1983;45:175–6.
37. Caroffino DC, Sutton TM, Daugherty DJ. Assessment of the vertical distribution of larval lake sturgeon drift in the Peshtigo River, Wisconsin, USA. J Appl Ichthyol. 2009;25:14–7.
38. Roseman EF, Manny B, Boase J, Child M, Kennedy G, Craig J, Soper K, Drouin R. Lake sturgeon response to a spawning reef constructed in the Detroit river. J Appl Ichthyol. 2011;27:66–76.
39. Bouckaert EK, Auer NA, Roseman EF, Boase J. Verifying success of artificial spawning reefs in the St. Clair-Detroit River System for lake sturgeon (*Acipenser fulvescens* Rafinesque, 1817). J Appl Ichthyol. 2014;30:1393–401.
40. Ortlepp J, Mürle U. Effects of experimental flooding on brown trout (*Salmo trutta fario* L.): the River Spöl, Swiss National Park. Aquat Sci. 2003;65(3):232–8.
41. Bilotta GS, Milner AM, Boyd IL. Quality assessment tools for evidence from environmental science. Environ Evid. 2014;3(1):1.
42. Rohatgi A. WebPlotDigitizer: HTML5 based online tool to extract numerical data from plot images. Version 4.1. 2015. https://automeris.io/WebPlotDigitizer/. Accessed 10 Jan 2018.
43. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Effect sizes based on means. In: Borenstein M, Hedges LV, Higgins JPT, Rothstein HR, editors. Introduction to meta-analysis. London: Wiley; 2009. p. 21–32.

44. R Development Core Team. R: a language and environment for statistical computing. Vienna: R foundation for Statistical Computing; 2017. http://www.R-project.org. Accessed 25 Jan 2018.
45. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1–48.
46. Hedges LV, Olkin I. Statistical methods for meta-analysis. New York City: Academic Press; 1985.
47. Light RJ, Pillemer DB. Summing up: the science of reviewing research. Cambridge: Harvard University Press; 1984.
48. Zwetsloot PP, Van Der Naald M, Sena ES, Howells DW, IntHout J, De Groot JA, et al. Standardized mean differences cause funnel plot distortion in publication bias assessments. eLife. 2017;6:e24260.
49. Rosenberg MS. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. Evolution. 2005;59:464–8.
50. Rosenthal R. The, "file drawer problem" and tolerance for null results. Psychol Bull. 1979;86:638–41.
51. Rosenthal R. Meta-analytic procedures for social research. London: Sage; 1991.
52. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics. New York: Springer; 2005.
53. Zeug SC, Sellheim K, Watry C, Rook B, Hannon J, Zimmerman J, Cox D, Merz J. Gravel augmentation increases spawning utilization by anadromous salmonids: a case study from California, USA. River Res Appl. 2014;30(6):707–18.
54. Pedersen ML, Kristensen EA, Kronvang B, Thodsen H. Ecological effects of re-introduction of salmonid spawning gravel in lowland Danish streams. River Res Appl. 2009;25(5):626–38.
55. West CJ, Mason JC. Evaluation of sockeye salmon (*Oncorhynchus nerka*) production from the Babine Lake Development Project. In. Smith HD, Margolis L, Wood CC, editors. Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Department of Fisheries and Oceans; 1987. pp. 176–90.
56. Benoit J, Legault M. Assessment of the feasibility of preventing reproduction of lake charr, *Salvelinus namaycush*, in shallow areas of reservoirs affected by drawdowns. Environ Biol Fish. 2002;64:303–11.
57. Scruton DA, Clarke KD, Roberge MM, Kelly JF, Dawe MB. A case study of habitat compensation to ameliorate impacts of hydroelectric development: effectiveness of re-watering and habitat enhancement of an intermittent flood overflow channel. J Fish Biol. 2005;67:244–60.
58. Marsden JE, Perkins DL, Krueger CC. Recognition of spawning areas by lake trout: deposition and survival of eggs on small, man-made rock piles. J Great Lakes Res. 1995;21:330–6.
59. Sternecker K, Wild R, Geist J. Effects of substratum restoration on salmonid habitat quality in a subalpine stream. Environ Biol Fish. 2013;96(12):1341–51.
60. Kondolf GM. Five elements for effective evaluation of stream restoration. Restor Ecol. 1995;3:133–6.
61. Kondolf GM, Micheli ER. Evaluating stream restoration projects. Environ Manag. 1995;19:1–15.
62. Rubin Z, Konfold GM, Rios-Touma B. Evaluating stream restoration projects: what do we learn from monitoring? Water. 2017;9(3):174.
63. Thompson B. Effect sizes, confidence intervals, and confidence intervals for effect sizes. Psychol Sch. 2007;45:423–32.
64. Roni P, Beechie TJ, Bilby RE, Leonetti FE, Pollock MM, Pess GR. A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in Pacific Northwest watersheds. N Am J Fish Manag. 2002;22(1):1–20.
65. Chernick MR. When bootstrapping fails along with remedies for failures. In: Chernick MR, editor. Bootstrap methods: a guide for practitioners and researchers. Hoboken: Wiley; 2007. p. 172–87.
66. Pek J, Wong ACM, Wong OCY. Confidence intervals for the mean of non-normal distribution: transform or not to transform. Open J Stat. 2017;7:405–21.
67. Kondolf GM. Assessing salmonid spawning gravel quality. Trans Am Fish Soc. 2000;129(1):262–81.
68. Stewart GB, Bayliss HR, Showler DA, Pullin AS, Sutherland WJ. Does the use of in-stream structures and woody debris increase the abundance of salmonids? CEE review 05-006 (SR12). Collaboration for Environmental

Taylor *et al. Environ Evid*   (2019) 8:19

Page 31 of 31

Evidence; 2006. http://www.environmentalevidence.org/SR12.html. Accessed 1 Feb 2018.

69. Soulsby C, Youngson AF, Moir HJ, Malcolm IA. Fine sediment influence on salmonid spawning habitat in a lowland agricultural stream: a preliminary assessment. Sci Total Environ. 2001;265(1–3):295–307.

70. Greig S, Sear D, Carling P. A field-based assessment of oxygen supply to incubating Atlantic salmon (*Salmo salar*) embryos. Hydrol Process. 2007;21(22):3087–100.

71. Fenkes M, Fitzpatrick JL, Ozolina K, Shiels HA, Nudds RL. Sperm in hot water: direct and indirect thermal challenges interact to impact on brown trout sperm quality. J Exp Biol. 2017;220:2513–20.

72. Nilsson C, Sarneel JM, Palm D, Gardeström J, Pilotto F, Polvi LE, et al. How do biota respond to additional physical restoration of restored streams? Ecosystems. 2017;20(1):144–62.

73. Albanese B, Angermeier PL, Peterson JT. Does mobility explain variation in colonisation and population recovery among stream fishes? Freshw Biol. 2009;54(7):1444–60.

74. Peel M, Finlayson B, McMahon T. Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci. 2007;11:1633–44.

75. Bond NR, Lake PS. Local habitat restoration in streams: constraints on the effectiveness of restoration for stream biota. Ecol Manag Restor. 2003;4(3):193–8.

76. Groot C, Margolis L, editors. Pacific salmon life histories. Vancouver: UBC Press; 1991.

77. Behnke R. Trout and salmon of North America. New York: Simon and Schuster; 2010.

78. Shetter DS, Clark OH, Hazzard AB. The effects of deflectors in a section of a Michigan trout stream. Trans Am Fish Soc. 1949;76:248–78.

79. Jester DB, McKirdy HJ. Evaluation of trout stream improvement in New Mexico. In: Proceedings of the annual conference of western association of state game and fish commissioners, vol. 46. 1966; p. 316–33.

80. Pickering H, Whitmarsh D. Artificial reefs and fisheries exploitation: a review of the 'attraction versus production' debate, the influence of design and its significance for policy. Fish Res. 1997;31(1–2):39–59.

81. Roni P. Does river restoration increase fish abundance and survival or concentrate fish? The effects of project scale, location, and fish life history. Fisheries. 2019;44(1):7–19.

82. Elliott JM. The critical-period concept for juvenile survival and its relevance for population regulation in young sea trout, *Salmo trutta*. J Fish Biol. 1989;35(sA):91–8.

83. Beamish RJ, Mahnken CA. Critical size and period hypothesis to explain natural regulation of salmon abundance and the linkage to climate and climate change. Prog Oceanogr. 2001;49(1–4):423–37.

84. Minns CK, Kelso JR, Randall RG. Detecting the response of fish to habitat alterations in freshwater ecosystems. Can J Fish Aquat Sci. 1996;53(S1):403–14.

85. Chapman MG, Underwood AJ. The need for a practical scientific protocol to measure successful restoration. Wetl Aust J. 2010;19(1):28–49.

86. Smokorowski KE, Bradford MJ, Clarke KD, Clément M, Gregory RS, and Randall RG. 2015. Assessing the effectiveness of habitat offset activities in Canada: monitoring design and metrics. Canadian Technical Report of Fisheries and Aquatic Sciences.

87. Smokorowski KE, Randall RG. Cautions on using the before-after-control-impact design in environmental effects monitoring programs. FACETS. 2017;2:212–32.

88. Harper DJ, Quigley JT. No net loss of fish habitat: a review and analysis of habitat compensation in Canada. Environ Manag. 2005;36(3):343–55.

89. Bernes C, Macura B, Jonsson BG, Junninen K, Müller J, Sandström J, Lõhmus A, Macdonald E. Manipulating ungulate herbivory in temperate and boreal forests: effects on vegetation and invertebrates. A systematic review. Environ Evid. 2018;17:13.

90. Eales J, Haddaway NR, Bernes C, Cooke SJ, Jonsson B-G, Kouki J, Petrokofsky G, Taylor JJ. What is the effect of prescribed burning in temperate and boreal forest on biodiversity, beyond tree regeneration, pyrophilous and saproxylic species? A systematic review. Environ Evid. 2018;7:19.

91. Souchon Y, Sabaton C, Deibel R, Reiser D, Kershner J, Gard M, Katopodis C, Leonard P, Poff NL, Miller WJ, Lamb BL. Detecting biological responses to flow management: missed opportunities; future directions. River Res Appl. 2008;24(5):506–18.

92. O'Neal JS, Roni P, Crawford B, Ritchie A, Shelly A. Comparing stream restoration project effectiveness using a programmatic evaluation of salmonid habitat and fish response. N Am J Fish Manag. 2016;36(3):681–703.

93. McCarthy MA, Possingham HP. Active adaptive management for conservation. Conserv Biol. 2007;21(4):956–63.

94. Haddaway N, Macura B, Whaley P, Pullin A. ROSES for systematic review reports. Version 1.0. 2017. https://doi.org/10.6084/m9.figshare.5897272.

Publisher's Note