How do changes in flow magnitude due to hydropower operations affect fish abundance and biomass in temperate regions? A systematic review

by

Meagan Harper

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Department of Biology

Carleton University

Ottawa, Ontario

March 2021

Abstract

Altering the natural flow regime (an essential component of healthy fluvial systems) through hydropower operations may negatively impact freshwater fish populations. This systematic review investigates how changes in flow magnitude due to hydropower operations influence fish abundance and biomass. Following the guidelines of the Collaboration for Environmental Evidence, I examined all available evidence identified by a recent systematic map and search update. Of 103 articles (133 studies) included for data extraction and critical appraisal, 58 studies were included for quantitative meta-analysis. Fish abundance and biomass had variable responses to flow magnitude changes; overall mean effect sizes ranged from positive to negative, differing by study design and taxa. No consistent patterns in fish abundance and biomass responses to alterations in flow magnitude were identified and fish responses were context dependent. Responses may not be generalizable across systems impacted by hydropower where specific features of the system may be highly influential.

Keywords: Dam, Discharge, Evidence synthesis, Fish biomass, Fish density, Flow modification, Hydropower

Acknowledgements

I would like to thank my advisors, Dr. Steven Cooke, Dr. Joseph Bennett and Dr. Trina Rytwinski for their support throughout my degree (and a global pandemic). I am also grateful to Dr. Karen Smokorowski and Dr. Frances Pick, to the members of the Canadian Centre for Evidence-Based Conservation for their collaboration at many stages of this project, and for the advice and guidance provided by other members of the Advisory Team: Keith Clarke, Thomas Pratt, Neil Fisher, Alf Leake, Julian Olden and Daniel Boisclair. Additional thanks go to the many individuals who answered the call for literature and provided access to reports and other sources of grey literature to inform a comprehensive systematic review. I would also like to specifically thank the Carleton Library staff who worked exceptionally hard to find electronic copies of physical documents not available to the public due to COVID-19 restrictions early in the pandemic. Special thanks to Dr. Mary Freeman for providing raw data from a previous work. This project would not have been possible without the many people involved in the original systematic map which generated the knowledge cluster this review is based on. This review was primarily supported by Fisheries and Oceans Canada. Additional support was provided by the Natural Science and Engineering Research Council of Canada and Carleton University. Finally, thank you to my family who provided me with a fabulous place to work throughout the pandemic…and many cups of tea.

Thesis Format and Co-authorship

This thesis consists of a single systematic review without chapters. It contains my own research but was conducted in a collaborative manner with members of a review team from the Canadian Centre for Evidence-based Conservation and an Advisory Team made up of stakeholders, academic scientists from Canada and USA, staff from DFO, specifically the Fish and Fish Habitat Protection Program (FFHPP), and the Science Branch, as well staff from hydropower industry. The Advisory Team was consulted throughout the review process.

Coauthors:

Rytwinski T, Taylor JJ, Bennett JR, Smokorowski KE, Clarke K, Pratt T, Fisher N, Leake A, Olden J, Boisclair D, Cooke SJ

Table of Contents

List of Tables

List of Figures

List of Appendices

.

1. Background

Humans act as a dominant force on Earth, altering ecosystems in diverse, often detrimental ways (Vitousek et al. 1997). Alterations of ecosystems span terrestrial and aquatic systems but are exceptionally apparent in fluvial systems (i.e., rivers and streams) where dams, constructed to generate hydroelectricity, have dramatically altered ecosystem structure and function (e.g., Nilsson et al. 2005; Winemiller et al. 2016). Globally, there are currently over 8,600 hydropower dams higher than 15 m and future hydropower projects, in development, are expected to double hydropower generation capacity over the next 50 years (Zarfl et al. 2015). Consequently, improving understanding of the impacts and ecosystem alterations associated with these projects is becoming increasingly essential for effective management of both hydropower production and important ecosystem components.

Maintaining the ecological characteristics of fluvial systems altered by the demands of hydroelectric power production (HPP) and related operations requires the careful management of flow regime components including magnitude, duration, frequency, timing and rate of change. Natural flow regimes have regulated both geological and biological components of natural waterbodies through time (Poff et al. 1997; Bunn and Arthington 2002; Naiman et al. 2008) and aquatic biota have evolved and adapted to the specific dynamics of their environment (Poff et al. 1997; Bunn and Arthington 2002; Lytle and Poff 2004). One of the critical components of a flow regime is flow magnitude (also called discharge), the measure of the volume of water passing a fixed location per unit of time (e.g., m^3/s) (Poff et al. 1997). Alterations to this component of the flow regime can occur in four ways (Fig. 1): (i) changes to maximum flow magnitude (also called peak flow); (ii) changes to minimum flow magnitude (also called base flow, but here referring only to base flow controlled by HPP); (iii) changes to the average flow magnitude (i.e.,

the average yearly flow magnitude or discharge); and (iv) changes in short-term variation (i.e., changes in magnitude that occur over a short period of time, such as 24 hours). Alterations to flow magnitude associated with HPP result in either increases or decreases in flow magnitude and can disrupt natural processes resulting in a variety of environmental and species responses (Bunn and Arthington 2002). Understanding how these alterations impact fluvial systems is important for water resource and fisheries management.

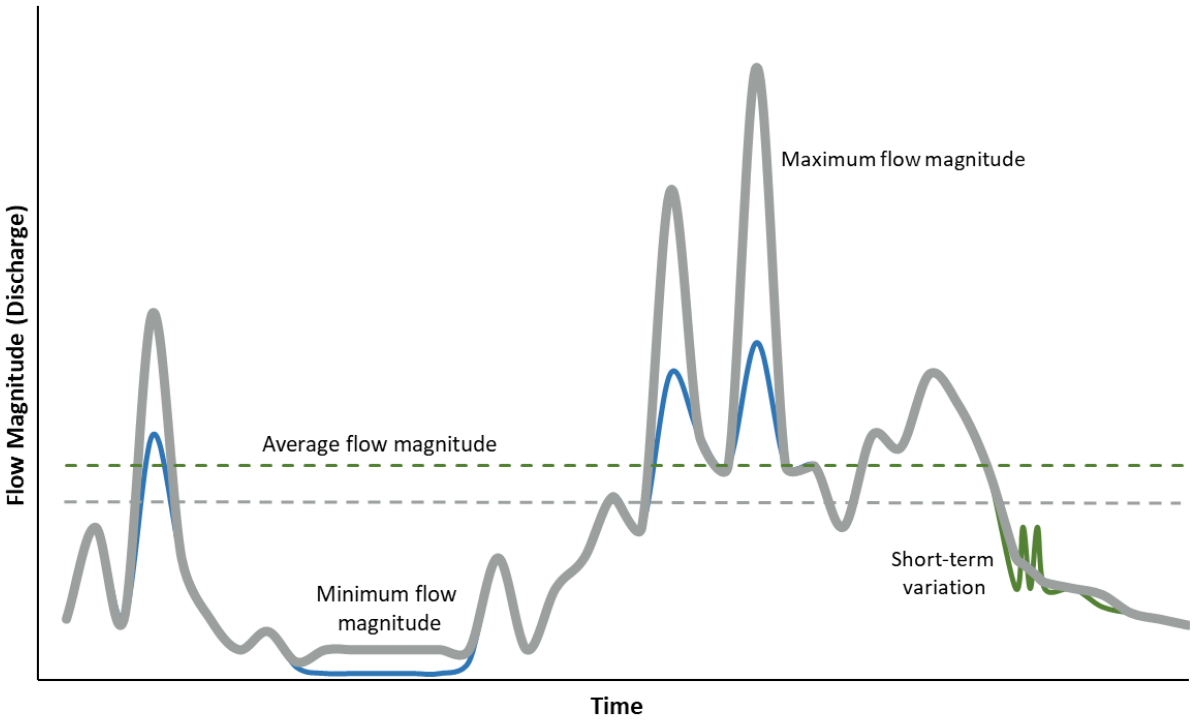

Fig. 1. Generalized hydrograph of a natural flow regime (grey) demonstrating potential changes in flow magnitude including increases (green) and decreases (blue) for: (i) maximum flow magnitude (also called peak flow); (ii) minimum flow magnitude (also called base flow); (iii) average flow magnitude (dotted lines; also called average discharge); and (iv) short-term variation (changes in magnitude that occur over a short period of time, usually 24 hours). Changes to any element of flow magnitude can be either increases or decreases.

The effects of HPP on fish living in or traveling through fluvial systems can include alterations to fish abundance (number of individuals, often quantified in terms of fish per area or catch) and biomass (total mass of individuals by area or volume) which may decrease or increase

in response to these changes in flow. Studies have shown that community abundance and biomass can differ between areas that are regulated by HPP facilities and those that are not (e.g., Kinsolving and Bain, 1993; Guénard et al., 2016). Fish abundance has also been found to decrease after negative alterations in flow magnitude (Poff and Zimmerman 2010; Webb et al. 2013), but conversely, may increase after the establishment of positive changes such as increases in minimum-flow releases (e.g., fluvial specialists increased in density compared to before-minimum conditions; Travnichek et al. 1995). Additionally, systems with flow magnitude that more closely mimics natural flow may have greater abundance than those with more altered flow magnitudes (e.g., systems with a legislated minimum flow magnitude compared to those without; Göthe et al. 2019). Establishment of better managed flow magnitude can therefore have positive impacts on fish species. These studies indicate that fish responses may be dependent on the type of HPP facility, the type of "designer" flow regime or near-natural flow regime (Acreman et al. 2014), and the magnitude of alteration.

Almost half of all rivers globally are altered by river regulation or fragmentation (Grill et al. 2019), and hydropower dams are a major contributor to these alterations. Hydroelectric power production is recognized as a continuing threat to freshwater species (Reid et al. 2019), especially with the increasing construction of both large and small HPP facilities (Zarfl et al. 2015; Couto and Olden 2018). With hydropower expected to continue to represent a large portion of many countries' energy portfolios (IHA, 2020) and to be increasingly utilized with the global move towards greener energy sources (Zarfl et al. 2015; U.S. Department of Energy 2016; National Energy Board 2018), understanding how alterations of specific flow components impact fish responses is essential. The effective management of flow regimes to provide flow characteristics that support both fish productivity and energy production in systems affected by

HPP, requires a better understanding of how fish respond to flow component alterations at hydroelectric dams, and may even require a re-evaluation of how modified river flows are designed (e.g., Soininen et al., 2019; Tonkin et al., 2019).

Available evidence syntheses on the impacts of HPP on fish often focus on the effects of passage on behaviour, injury and/or mortality of fish due to HPP facilities (Coutant and Whitney 2000; Pracheil et al. 2016; Algera et al. 2020), or on the alteration in abundance and diversity of fish populations resulting from specific types of hydropower operation (i.e., hydropeaking; Melcher et al., 2016) or design (i.e., impoundments; Turgeon et al., 2019). While reviews on ecological responses to altered flows have been done in the past (Murchie et al. 2008; Webb et al. 2013; Gillespie et al. 2015) and a recent narrative overview considered the various aspects of ecohydrology river alterations on fish (Boavida et al. 2020), there remains a need to update our understanding of specific fish-flow interactions using robust systematic review techniques. Additionally, there is a need to reduce the uncertainty surrounding how fish respond to alterations in specific flow components such as flow magnitude (Rytwinski et al. 2020).

Systematic reviews are valuable tools of evidence synthesis developed to answer a specific question in an unbiased way and to increase understanding of the impacts of anthropogenic stressors (Pullin and Knight 2009). They are increasingly used to inform decision-making in environmental management and policy (Cook et al. 2013; Bilotta et al. 2014; Collins et al. 2019), because their emphasis on objectivity, transparency, rigour and repeatability, helps to provide certainty for informed decisions (Cooke et al. 2016). A unique feature of this type of review (which collates and synthesizes all available, relevant evidence on a topic) is the use of critical appraisal to test the methodological reliability and relevance of evidence to the question of interest, and to determine the confidence that can be placed in study findings (Pullin and

Stewart 2006; CEE 2018). The guidelines provided by the Collaboration for Environmental Evidence (CEE) help ensure that evidence syntheses are conducted in a rigorous, transparent and repeatable manner (CEE, 2018). A systematic review of how flow components such as magnitude, altered by HPP, affects fish abundance and biomass would help support effective flow management decision making. Here, we[1] use a systematic review approach, including meta-analysis, to evaluate the existing literature base to assess the consequences of alterations to flow magnitude by HPP on fish abundance and biomass, and to identify to what extent factors such as dam size, operational regime, direction of flow magnitude alteration, and life history characteristics influence the response of fish abundance to these changes.

2. Identification of review topic and stakeholder engagement

At the request of Canadian stakeholders [i.e., Fisheries and Oceans Canada (DFO)], a systematic map was recently conducted (Rytwinski et al. 2020) to provide a summary of the existing literature base on the impacts of flow regime changes on direct outcomes of freshwater and estuarine fish productivity (i.e., the map described the quantity and key characteristics of the available evidence, identified evidence clusters and knowledge gaps, but did not synthesize results). A total of 1368 relevant studies describing a variety of flow regime alterations and fish productivity responses, were identified. The map focused on global temperate regions and followed the CEE guidelines for systematic mapping (CEE 2018).

[1] This thesis contains my own research but was conducted in a collaborative manner with members of the review team and the Advisory Team. I accept full ownership of, and responsibility for, the content but have elected to use *we* and *our* throughout to reflect the collaborative nature of this work.

From the systematic map, 11 potential topic clusters were identified as areas that had sufficient coverage to allow for systematic reviewing. The subtopic "the effect of alterations to flow magnitude due to hydroelectric power production on fish abundance" was identified as a candidate for full systematic reviewing based on the presence of sufficient evidence and interest from Canadian stakeholders. Although not included in this review, the subtopic "the effects of alterations to flow magnitude due to dams with no hydroelectric facilities on fish abundance" (46 studies) was also identified as a candidate for future systematic reviewing. Canadian stakeholders acknowledge that a comparison of the relative effects of dams with and without hydroelectric facilities on fish abundance, diversity, and richness deserves future attention; however, such comparison was beyond the scope of this review.

An Advisory Team made up of stakeholders and experts including academic scientists from Canada and USA (four members), staff from DFO, specifically the Fish and Fish Habitat Protection Program (FFHPP) (one member), and the Science Branch (three members), as well staff from hydropower industry (one member) was established and consulted during this review process. The Advisory Team was consulted in the development of the inclusion criteria for article screening and data extraction strategy and participated throughout the course of this systematic review.

3. Objective of the review

The objective of this systematic review was to clarify, from existing literature, how fish abundance and biomass are impacted by alterations of flow magnitude due to hydroelectric power production (or related operations) to better inform decisions in water resource and fisheries management downstream of these facilities.

3.1 Primary question

How do changes in flow magnitude due to hydroelectric power production affect fish abundance and biomass in temperate regions?

3.2 Components of the primary question

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

Subject (population) – freshwater and estuarine fish in temperate regions

Intervention/exposure – changes to (or manipulation of) flow magnitude due to hydroelectric power production

Comparator – no intervention or alternate levels of intervention

Outcomes – measures of changes in abundance (e.g., abundance, density, catch per unit effort), and biomass (e.g., biomass, yield)

4. Methods

This review followed the CEE guidelines and standards for systematic reviews (CEE 2018) and conformed to ROSES (RepOrting standards for Systematic Evidence Syntheses in environmental research) reporting standards (Haddaway et al. 2018) (see Appendix 1). The methods of this review follow those published in an *a priori* systematic review protocol (Appendix 2; Harper et al. 2020). We summarize the methods here and describe any deviations from the protocol made during the conduct of the review, below.

4.1 Searching for articles

4.1.1 Selection of studies identified in the systematic map

Much of the evidence for this systematic review was identified by the recent Rytwinski et al. (2020) systematic map on fish productivity and flow alteration. The systematic map searched for

commercially published and grey literature using six bibliographic databases (performed in July 2017), one search engine (July 2017) and 29 specialist websites (Feb. 2017). In addition, 297 relevant reviews and all accepted articles were hand-searched for relevant titles not found using the search strategy. Calls for evidence to target grey literature were also issued through relevant mailing lists, social media and by the Advisory Team members to their networks and colleagues. A total of 1368 relevant studies were identified by this map, with 74 considering flow magnitude alterations and fish abundance and 24 considering fish biomass metrics. All potentially relevant studies identified by the systematic map were initially included for this review at the data extraction stage. During data extraction, these studies were further screened on the specific eligibility criteria of this review.

4.1.2 Search update

Search terms and language

Search terms used in the systematic map prior to this systematic review that identified studies considering the impacts of alteration to any component of flow on fish productivity can be found in Appendix 3 (Table S10).

An updated search was conducted on a subset of the search terms used in the systematic map (Table 1). These terms were used to query bibliographic databases (see section "*Publication Databases*" below) and the search engine, Google Scholar (see section "*Search engines*" below"). The updated search covered literature published from 2017 through 2019. Search terms were limited to English language due to project resource restrictions; however, no language, geographic, or document type restrictions were applied during the search. The search string was modified depending on the search functionality of different databases or the search engine (see

Appendix 3). Full details on search settings and subscriptions used to access articles can be found in Appendix 3.

Table 1. Search string used to update searches from 2017 through 2019.

Component	Search string
Population terms	((Fish*) AND (“Fresh water” OR Freshwater OR Stream$ OR Water$ OR River$ OR Fluvial OR Estuar* OR Reservoir$ OR Impoundment$ OR "Hydro electric*" OR Hydroelectric* OR "Hydro dam*" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam$))
	AND
Intervention/exposure terms	(Flow* OR Discharg*)
	AND
Outcome terms	(Productivity OR Biomass OR Abundance$ OR Densit* OR Yield$ OR “Ecological response” OR “Ecosystem response” OR “Biotic response”)
	NOT
Exclusionary terms	(Mining OR "Mine site" OR Aquaculture OR "Wastewater treatment" OR Carbon)

Publication databases

To ensure sufficient coverage and specificity during searching, the principle search system ISI Web of Science Core Collection was used (Gusenbauer and Haddaway, 2019) and an additional five bibliographic databases were also accessed. All databases (listed below) were originally searched in the map, and search updates occurred in November-December 2019 using Carleton University’s institutional subscriptions:

1. ISI Web of Science Core Collection
2. ProQuest Dissertation & Theses Global
3. Scopus
4. Federal Science Library (Canada)
5. Science.gov
6. AGRICOLA (Agricultural Research Database)

Search engines

To supplement our principle searches and identify potentially useful documents not already found by database searches, the same search engine originally used in the systematic map, Google Scholar, was searched January 2020 (first 500 hits sorted by relevance). Potentially relevant documents were recorded and included to be screened for appropriate fit with the review question. Customized search strings were used due to limited search capability of the search engine (see Appendix 3).

Specialist websites

Twenty-nine specialist organization websites were searched in the systematic map using abbreviated search terms (see Rytwinski et al. 2017). A search update was not conducted for these websites because it is often not possible to specifically filter by date using the built-in search functions of these websites.

Supplemental Searches

Reference sections of accepted articles and 110 relevant reviews (2 relevant reviews were removed as duplicates) were hand-searched to evaluate relevant titles, published from 2017 forward, not identified using the search update strategy (see Appendix 4 for a list of relevant reviews). Stakeholders were consulted for advice for new sources of information. We also issued a call for evidence to experts and practitioners in the field to target grey literature through relevant mailing lists [Canadian Conference for Fisheries Research, American Fisheries Society, WaterPower Canada (formerly Canadian Hydropower Association), Canadian Electricity Association, Ontario Women Anglers, the Mactaquac Aquatic Ecosystem Study, Instream Flow Council] and through social media (e.g., Twitter) in December 2019. The call for evidence was also distributed by the Advisory Team to relevant networks and colleagues. If experts and

practitioners suggested websites or databases not already captured during the systematic mapping exercise or search update, these sites were either hand-searched (and articles included for screening at full-text) or, where possible, searched using built-in functions and modified keywords (see Appendix 3). These included:

1. ARLIS – Alaska Resources Library and Information Services
2. FERC Online eLibrary – US Federal Energy Regulatory Commission eLibrary

In cases where articles were found using built-in search functions, articles were included for eligibility screening at the title and abstract stage (see Appendix 3). To increase the chance of capturing previously missed unpublished information from expert and practitioner recommendations, no date restrictions were applied. Additionally, in one case, experts and practitioners suggested a specific set of documents from a website already accessed during the systematic map (BC Hydro, Water Use Plans). Water Use Plan projects were screened at title and abstract (e.g., we removed projects on ineligible topics such as archeology). Articles within applicable projects were accessed and checked against the results of the systematic map for duplicates. To ensure all relevant articles were captured, articles from applicable projects not previously identified were included for eligibility screening at title and abstract, and full text screening.

Estimating comprehensiveness of the search

For this review, we did not repeat tests for comprehensiveness originally performed in the systematic map (i.e., the search results were checked against a benchmark list of 13 relevant papers provided by the advisory team to ensure all articles were captured using the search strategy). Because the review followed the same basic search strategy and used a search string similar to the systematic map (Rytwinski et al. 2017), further comprehensiveness checks were

not necessary. The majority of articles included as relevant in the systematic map (using a broader eligibility criteria than this review) were identified through databases and search engines (88%), reference sections of reviews and included articles, or through calls for evidence (9%), with only 3% being identified through website searches. We therefore considered it sufficient to base the search update on the same databases and search engine used in the systematic map, complemented with the supplemental searches described immediately above. More specifically, we increased the likelihood of capturing relevant literature not identified from our search strategy by screening bibliographies of: (1) 110 individual relevant reviews identified at title and abstract or full text; (2) accepted articles. We searched these reference lists until the reviewer (MH) deemed that the number of relevant returns had significantly decreased.

4.2 Search record databases

Once all searches were complete and references were compiled, individual databases were exported to EPPI-reviewer (eppi.io.ac.uk/eppireviewer4) as one database. Prior to screening, duplicates were identified using the duplicate checking function in EPPI Reviewer and then manually removed by one reviewer (MH). All references, regardless of their perceived relevance to the systematic review, were included in the database. Results of supplemental searches were compiled in MS Excel and screened separately. Duplicate checks between the EPPI-reviewer database and supplementary searches were conducted by one reviewer (MH). When duplicates were missed during any previous stage, they were removed at later stages in the review.

4.3 Article screening and study eligibility criteria

4.3.1 Screening process

Articles found by database searches and the search engine (including those suggested during calls for literature) were screened at two stages: (i) title and abstract, and (ii) full-text. Other

articles found through supplemental searches were screened at full text. No articles found through supplemental searches were included in consistency checks. Prior to screening all articles, a consistency check was done at the title and abstract stage where two reviewers (JJT and MH) screened 181/1810 articles [10% of the articles included in the EPPI Reviewer (which did not include evidence items found through supplemental searches or literature identified by the systematic map)]. The reviewers agreed on 93.34% of the articles (kappa = 0.59; moderate agreement). Any disagreements between reviewers were discussed and the inclusion criteria clarified, before moving forward. Following consistency checks, articles were screened by one reviewer (MH). Reviewers did not screen (at title and abstract or full-text) any article to which they were an author. Attempts were made to retrieve full-texts of all articles included at title and abstract screening using the Carleton University library subscriptions or interlibrary loans. Authors of unpublished references or works that were unobtainable through library licenses or interlibrary loans were contacted to gain access to electronic copies. It was not possible to request physical copies of articles not available in electronic form due to COVID-19 public health restrictions at time of searching, which we acknowledge as a potential bias (Kadykalo et al. 2021).

A consistency check was also done at full-text screening with 12/122 articles [10% of articles included in EPPI Reviewer (which did not include items found through supplemental searches or literature identified by the systematic map)]. Reviewers (TR and MH) agreed on 83.33% of articles (kappa = 0.57; moderate agreement). Upon discussion, an inconsistency due to a missed detail in one article was resolved. Since this missed detail did not require a different application of the eligibility criteria, final agreement was actually 91.67% (kappa = 0.75; substantial agreement) and full-text screening proceeded. Full-text screening was conducted by a

single reviewer (MH). A list of all articles excluded at full-text screening, with reasons for exclusion, is provided in Appendix 4. Articles identified from the systematic mapping exercise were screened for eligibility at the data extraction stage by a single reviewer (MH). Any article excluded during data extraction, along with reasons, is included in the full list of excluded articles (Appendix 4).

4.4 Eligibility criteria

All articles had to meet the following criteria [following a PICO (Population, Intervention, Comparator, Outcome) framework], modified from the systematic map, to be included in the review.

4.4.1 Eligible populations

Relevant subjects included any resident (i.e., non-migratory) or migratory fish, including diadromous species (i.e., fish that migrate between fresh and salt water), in North (23.5°N – 66.5°N) or South (23.5°S – 66.5°S) temperate regions. Any life stage was considered. Populations could include those that were once stocked (but no longer being actively stocked) or invasive and that are established in the waterbody. Only articles considering fish species in freshwater or estuarine fluvial (i.e., water moving via gravity) ecosystems impacted by HPP systems (such as lakes, rivers and streams), were included.

4.4.2 Eligible interventions/exposures

Articles that described a change in, or modification to, the magnitude of downstream flow as a direct result of HPP facilities were included (whether flow varied directly as a result of hydropower production or due to related operational changes such as spilling for safety or water management). Magnitude, defined here as the amount of water moving past a fixed location per unit time, can be a direct measure of discharge, or expressed as a relative or absolute change

(Poff et al. 1997). Only fluvial effects of changes to flow magnitude were considered. Changes in, or modifications to, flow magnitude upstream of HPP facilities were not considered. Articles considering other flow component alterations (i.e., frequency, duration, timing or rate of change) were excluded if magnitude was not also considered. Relevant causes of flow alteration included HPP facilities where water moved via gravity (i.e., hydropeaking, impoundment or diversion/run-of-river) or by active pumping. Operations that may have impacted flow magnitude, but that were not related to HPP, were excluded. These included but were not limited to: (i) nuclear facilities; (ii) dams without hydropower; (iii) hydrokinetic systems (i.e., energy from waves/currents); and (iv) water withdrawal/diversion systems not associated with HPP. Studies that considered environmental flow augmentation were included if they were associated with HPP facilities. Changes in flow magnitude due to other environmental alterations (i.e., land-use change) or natural causes (i.e., climate change or extreme weather events) without also including the impact of a HPP facility were excluded. Articles with flow magnitude changes due to natural causes were identified for an upcoming systematic review (Birnie-Gauvin et al. in review), but were not considered in this review. At the request of stakeholders, articles that did not specify a flow component [e.g., the study compared an unregulated stream or stream section to a regulated stream (i.e., regulated via a hydro dam)] or reported unspecified multiple components of flow but did not report the effects separately to isolate individual impacts of the flow components, were included.

4.4.3 Eligible comparators

Relevant comparators included: (i) similar sections of the same waterbody with no intervention (e.g., upstream conditions); (ii) separate but similar waterbodies with no intervention; (iii) *Before* intervention data within the same waterbody (i.e., pre-construction/modification/operation); (iv)

alternative levels of intervention on the same or different waterbody; and (v) controlled flume studies (note, no articles of this type were identified during the review process). When authors stated that the comparator was downstream of the HPP site, articles were excluded at the initial data extraction stage to determine a count of this type of article. Based on stakeholder feedback, we assumed that any site along the full distance of a river experiences the effects of hydropower modification upstream, but with a time delay in relation to upstream sites. Although authors sometimes reported a return to 'near normal' flows at downstream control sites, this is not considered to be comparable to control sites that never experience an impact from a HPP system. Therefore, we did not include studies with downstream controls, even if explicitly identified as a control by the authors. Additionally, if upstream comparator sampling sites (i.e., sites in the same waterbody and in upstream conditions) were mostly in free flowing sections of the river, but a minority of sampling sites were in a 'transition' zone between the free-flowing section of the river and a reservoir (i.e., reservoir tails), all sampling sites were considered as controls but this study design characteristic was acknowledged during study validity assessment (i.e., the study was assessed as having intervention and control sites that were moderately matched; see Table 2). If, however, comparator sites were primarily within reservoirs or reservoir tails and no free-flowing sites were considered, these sites could not be considered an upstream comparator and the article was excluded.

4.4.4 Eligible outcomes

Included articles considered outcomes that indicated the potential for a change in fish abundance (broadly defined to include fish biomass). Outcomes included those related to: (i) abundance: abundance (number of individuals), density (number of individuals per sampled area), catch per unit effort (CPUE), and presence/absence, and (ii) biomass: biomass and yield. Fish passage

studies that determined the number of fish passing a particular HPP system were included only if they also considered abundance measured below the HPP facility in relation to a change in flow magnitude (i.e., measured numbers or types of fish before and after a flow change below the HPP facility). Passage studies that reported changes in number of individuals above and below the HPP facility or a downstream barrier and used these counts as indicators of fish passage (i.e., the difference in number of fish above and below a natural barrier before and after a change in flow) were excluded as it was not possible to determine if this was a true change in population abundance or simply a change in the number of fish moving from one site to another. Articles were also excluded if they only considered other direct responses of fish productivity (e.g., growth, survival, migration) or evaluated indirect links between measured outcomes and altered flow (e.g., growth of aquatic plants) and potential responses of fish (e.g., diversity).

4.4.5 Eligible types of study designs

This review considered primary, field-based studies including quantification of fish abundance and biomass outcomes using *Before/After* (*BA*), *Control/Impact* (*CI*), *Before/After/Control/Impact* (*BACI*), *Reference Conditional Approach* (*RCA*), *Normal Range* (*NR*), or *Randomized Controlled Trials* (*RCT*; e.g., small in-field manipulations). Also considered were *CI* designs comparing two levels of intervention on different water bodies (*ALT-CI*) and *CI* designs using a gradient of intervention intensity that included a "zero-control" site (i.e., unimpacted site) (*CI-gradient*). *CI-gradient* studies were originally considered for inclusion but were either converted to: (i) *CI* designs with pseudoreplication, if there were subsamples taken in the same river, or (ii) multiple, but non-independent *CI* studies, if studies compared multiple independent rivers with different HPP impacts to a "zero-control" site. Studies were excluded if they used: (i) temporal trends looking at the relationship/correlation between fish

abundance or biomass and changes in magnitude across time without a 'true' *Before* intervention time period; (ii) spatial trends that do not include "zero-control" site: (a) across waterbodies [e.g., survey fish abundance in six different streams (i.e., of different morphology) and relate to flow magnitude]; or (b) within a waterbody [e.g., survey of fish abundance in different sections of the same stream that differ in morphology (e.g., riffle and run), or where downstream comparators were considered]; (iii) >1 *After*-treatment time periods but no change/modification to flow magnitude occurred across time periods [i.e., repeat visits with no *Before*-treatment; *After-only* (*A-only*)]; (iv) >1 impact sites but no change in flow magnitude across impact sites occurred [i.e., multiple impact sites but no control site or *Before*-treatment data; *Impact-only* (*I-only*)]; (v) a single point in time with no comparison to another site; or (vi) a single impact site with no *Before*-treatment data. Theoretical modeling, reviews and policy discussions were excluded.

4.4.6 Language

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

4.5 Study validity assessment

All studies found to be relevant to this review at the full-text screening stage underwent a study validity assessment using a critical appraisal tool, informed by previous tools (e.g., Macura et al., 2019; Martin et al., 2020), developed specifically for this review (Table 2). Each study (see definition in Table 3) was critically appraised for internal validity (i.e., susceptibility to bias) and study clarity using the predefined criteria outlined in Table 2. If a study contained more than one project (i.e., the project differed in terms of one or more components of the study validity; Table 2), each project received a separate validity rating and label (i.e., Enders et al. 2017 a/b indicates that there is one study with two projects within the Enders et al. 2017 article). For example,

temporal and spatial comparisons [i.e., whether the study compared samples temporally (*BA*), spatially (*CI*) or using a combination (*BACI*)] was an internal study validity criterion (Table 2). If one project within the study had a *BACI* comparison, while the other had a *BA* comparison, they received different internal validity responses for this criterion and therefore a different project label. The appraisal tool was made in consultation with the Advisory Team to ensure that it incorporated the components of a well-designed study.

The criteria of the appraisal tool are based on internal validity and evaluated the following: (i) study design [i.e., *BACI*, *RCT*, *BA*, *CI*, *ALT-CI*, *RCA* and *NR*, as well as deficient *BA* designs (DEF_BA) and incomplete *BACI*s (i.e., *BA* studies with *Before* data from a different site or studies missing certain components of a true *BACI* design; INCOM_BACI)], (ii) replication (no replication, pseudoreplication, or true replication), (iii) control and site matching (i.e., how well matched were intervention, comparator and sub-sample sites in terms of habitat, river section etc. at the outset of the study and during sampling), (iv) the intervention metric (quantitative or qualitative changes in magnitude or comparisons of unregulated to regulated systems), (v) confounders (other environmental or anthropogenic changes or other additional manipulations of flow components that occurred after the start of the study) and (vi) methodology (i.e., were different methodologies used between intervention and comparator sites and/or times). Each criterion was scored as 'Yes', 'Partially', 'No' or 'Unclear' based on the pre-defined criteria in Table 2. The study was given an overall 'Low' validity if it scored 'No' or 'Unclear' for any one or more of the criteria. If a study did not score 'No' for any criteria, but scored 'Partially' for one or more criteria, it received an overall 'Medium' validity score. Studies for which the answer was 'Yes' to all questions were classified as having 'High' validity. Each criterion received equal weight during study validity assessment. External validity (study

generalizability) was not directly assessed; instead, generalizability was captured during the screening stage, during data extraction or otherwise noted as a comment in the critical appraisal tool. In accordance with CEE guidelines (CEE 2018), reviewers would not have assessed study validity or conduct critical appraisal on studies for which they were an author; however, this situation never arose.

Study validity assessment took place at the same time as data extraction and was performed by a single reviewer (MH). A consistency check on the meta-data extraction/quantitative data extraction and study validity assessment was conducted by two reviewers (MH and TR) on 5/103 articles (5%) and quantitative data extraction was further tested on an additional three articles as extraction criteria were refined during data extraction. Meta-data extraction and study validity assessment were done by both reviewers and discrepancies were discussed (i.e., if one reviewer assigned 'Partially' for replication, while the other assigned 'Low', the reason for this discrepancy was discussed and clarification on what constitutes 'No' replication was included in critical assessment tool; see Table 2). When necessary, refinements to the meta-data extraction sheet and validity assessment tool were made to improve clarity of coding and the criteria. No study was excluded on the basis of study validity assessments; however, a sensitivity analysis was carried out to investigate the influence of study validity categories (see "Sensitivity analyses" below).

Table 2. Critical appraisal tool for study validity assessment. Additions made to the tool since the protocol are in italics.

Question/criterion	**Response to question**				**Type of Bias addressed**
	Yes	**Partially**	**No**	**Unclear**	
1. Did the study consist of both temporal and spatial comparisons?	BACI, RCT	INCOM-BACI, BA, CI, ALT-CI, RCA, NR	N/A as study is not eligible for inclusion based on inclusion criteria	*DEF-BA* or lacking sufficient information to judge	selection
2. Are experimental/observational units replicated?	≥2 independent experimental/observ-ational units (i.e., the level of replication at which the intervention was administered/the exposure experienced). *For BA designs, is there either within-year replication (fish abundance data available for at least 2 months/seasons post-intervention) or interannual replication (fish abundance data available for at least two years post-intervention).*	There were at least two experimental/observ-ational units but there is a lack of independence between these units (pseudoreplication). *N/A for BA designs.*	No replication (i.e., <2 independent experimental/ observational units), *or, for CI designs, the study had replication (either true or pseudo), but (sub)sample sites were too different to be treated as replicates (i.e., major differences in magnitude, morphology or habitat characteristics, and/or the presence of intersecting tributaries). For BA designs, fish abundance data was only available for a single post-intervention sample.*	Lacking sufficient information to judge	selection

Question/criterion	Response to question				Type of Bias addressed
	Yes	**Partially**	**No**	**Unclear**	
3. Are intervention and comparator sites well-matched at site selection and/or study initiation?	*Sites are well matched e.g.:* (1) Intervention and comparator sites are well-matched (i.e., similar physical characteristics), *(2) Sub-sample sites are well-matched within both intervention and comparator sites (e.g., comparator sub-sample sites are all within the free-flowing river section).* N/A for BA designs (i.e., intervention/comparator at same site).	*Sites are moderately matched e.g.:* (1) Intervention and comparator sites are moderately matched, *(2) sub-sample sites are moderately matched within intervention and/or comparator sites (e.g., comparator sub-sample sites are in free-flowing section and river/reservoir transition zones).* N/A for BA designs.	Intervention and comparator sites are poorly matched. N/A for BA designs.	Lacking sufficient information to judge	selection
4. Can the intervention be clearly interpreted?	It is clear that a change to flow magnitude has occurred and quantitative data on magnitude is reported	It is clear that a change to flow magnitude has occurred but either no quantitative data on magnitude is reported or the quantitative data is difficult to interpret (e.g., averaged across intervention and control sites within the same river)	N/A	The study compares an unregulated stream (or section of a stream) to a regulated stream (i.e., regulated via a hydro dam) or reports unspecified multiple components affecting flow (i.e., study does not report effects of components separately to isolate individual impacts of components)	selection, performance, reporting

Question/criterion	Response to question				Type of Bias addressed
	Yes	Partially	No	Unclear	
5. Was the study free of other potential confounders after sample selection/study initiation?	No or minimal confounding factors present, including e.g.: (1) no or minimal differences in environmental conditions between intervention and comparator sites and/or time periods (e.g., unplanned human alterations, floods, droughts, time-related trends), (2) no additional experimental manipulations of other flow regime components (e.g., flow frequency, duration) at the same time as magnitude alternations or, if present, are accounted for appropriately in analysis.	N/A	Confounding factors present that could have an impact on the outcome and these are not accounted for in analysis.	Lacking sufficient information to judge	performance
6. Did the study use similar sampling/ measurement method(s) between intervention and comparators?	Similar/consistent sampling/measurement methods are used between intervention and comparator sites and/or times (e.g., gear type, timing or size of sample areas)	Different/inconsistent sampling/ measurement methods are used between intervention and comparator sites and/or times (e.g., gear type, timing or size of sample areas)	N/A	Lacking sufficient information to judge	detection

Table 3. Definition of terms used throughout the systematic review.

Term	Definitions
Article	An independent publication (i.e., the primary source of relevant information). Can be from commercially published or grey literature sources. Used throughout the review.
Site	A specific hydroelectric facility (i.e., hydro dam) where observations or experiments were conducted and reported in one or more articles. Used throughout the review.
Study	An experiment or observation that was undertaken over a specific time period at particular sites reported as separate waterbodies that were not treated as replicates within a single article. Used throughout the review.
Project	Individual investigations within a study that differ with respect to ≥1 aspect of the study validity criteria (e.g., replication). Used in the review descriptive statistics and narrative review.
Case	Situationally defined in text/visual aids (e.g., separate counts of fish life stages) within an independent study. Used in review descriptive statistics and narrative review.
Dataset	(1) A single independent study from a single article; or (2) when a single independent study reported separate relevant comparisons for the same or different species and different: (a) operating conditions (e.g., different flow magnitudes/intensities, operational regime); (b) outcome categories (i.e., biomass, abundance, diversity, richness, or composition); (c) life stages for the same outcome category (e.g., the abundance of eggs for species X and the abundance of age-0 for species X) but otherwise with the same meta-data; (d) outcome metrics within a particular outcome category (i.e., abundance and density or CPUE; or biomass and yield) but otherwise the same meta-data; (e) sampling methods but otherwise with the same meta-data; (f) years and/or seasons post-treatment within a given outcome category (i.e., if for a given outcome category, multiple after time periods were monitored and reported separately for a *CI* study design or within-in year variation post-treatment for a *BA* design), and/or (g) sites downstream of a hydro dam within a single river sampled using a *BA* design but otherwise the same meta-data. The number of datasets was considered during quantitative analysis.

4.6 Data coding and extraction strategy

4.6.1 General data-extraction strategy

All articles identified from the search update that were included on the basis of full-text assessment underwent meta-data extraction. Articles identified as potentially relevant from the mapping exercise were further screened at this stage; if an article met the full eligibility criteria for this review, it underwent meta-data extraction. If an article was not deemed relevant, it was excluded from the review and recorded with the list of articles excluded at the full-text screening

stage, along with reasons (Appendix 4). Data extraction was conducted with a review-specific data extraction form (Appendix 5), following the general structure of our PICO framework. The following key variables of interest were developed through consultation with the advisory team: (i) bibliographical information; (ii) study location and details (e.g., geographic location, waterbody name and type); (iii) hydropower facility information (e.g., type, size, operational capacity); (iv) broad study objective; (v) study design and length; (vi) intervention/exposure (see Table 4 for definitions); (vii) comparator type; (viii) potential confounders (e.g., alterations to other flow components); (ix) outcome type; (x) sampling method(s); (xi) species [or species groups; common and Latin names crosschecked with FishBase (Froese and Pauly 2019) or Eschemeyer's Catalog of Fishes (Fricke et al. 2020)] and life stage(s) studied; and (xii) study validity assessment decisions. Coding within these key variables was based on codes previously developed during the systematic map (Rytwinski et al. 2017) and expanded through a partially iterative process as options were encountered during scoping and extraction.

Although we attempted to extract quantitative data on flow magnitude alterations (e.g., Δ change in flow magnitude) the complexity and variation of flow magnitude alterations in the eligible studies made extracting comparable quantitative intervention data impracticable due to limitations in time and resources. Based on stakeholder input, we generalized and characterized flow magnitude alteration by assigning categorical descriptors that capture both the primary change in flow magnitude element [i.e., changes to base flow, peak flow, average discharge or short-term variation (see Table 4 for definitions)] and the general direction of change (e.g., increase or decrease). Although this did not allow us to capture of the strength of change in flow magnitude, it did allow us to consider the impact of the direction of flow alteration and the type of flow magnitude change.

Table 4. Types of interventions, flow magnitude alterations considered (including elements and direction) and their definitions.

Intervention	Description	Code**
Alterations to flow magnitude elements due to hydropower	Any change in the amount of water moving past a fixed location per unit time (e.g., m^3/s), separated into four general elements:	
	1. Peak flow (reported as alterations in flood, peak, or high flow)	Peakflow
	2. Base flow (reported as alterations in base flow*, low flow or drought conditions)	Baseflow
	3. Average discharge (reported as alterations in total flow or mean flow)	AvgDischarge
	4. Short-term variation (reported as a change in magnitude that occurred over a period of hours or less than a day).	ShortVar
	5. Unspecified (no specified flow magnitude or direction of change [e.g., the study compares an unregulated stream (or section of a stream) to a regulated stream (i.e., regulated via a hydro dam)]; 2) reported as unspecified multiple flow magnitude elements and flow magnitude direction (i.e., do not report effects of elements separately to isolate individual impacts of flows magnitude elements)	UNSPEC
Increase flow magnitude	An increase in any flow magnitude element, either qualitatively or quantitatively reported by authors (e.g., increase in peak flow)	_Inc
Decrease flow magnitude	A decrease in any flow magnitude element, either qualitatively or quantitatively reported by authors (e.g., a change from 5 m^3/s base flow to 3 m^3/s base flow)	_Dec

*Base flow: Here used as a hydroelectrical operational term describing a minimum percentage of average flow, or the minimum allowable flow release from the hydropower facility, regardless of flow required for power generation needs. Our definition does not include baseflow from groundwater sources.
**Each intervention is a combination of flow alteration and direction (e.g., Peakflow_Inc or Baseflow_Dec).

Attempts were made to identify supplementary articles (i.e., articles that reported data that could be found elsewhere, that contained portions of information that could be used in combination with another more complete source, or articles that were yearly continuations of a previously established study) and combine them with the most comprehensive article (i.e., the primary source) during data extraction. Although separate laboratory experiments (flume studies) were originally considered potentially relevant, no laboratory experiments were identified during screening. When alternative *CI* studies occurred (*ALT-CI*), a comparator that was most similar to other *CI* studies with "zero-control" sites (i.e., natural or free-flowing rivers or stream sections)

was selected by the reviewer. In the one study where this occurred (i.e., Göthe et al. 2019), the systems altered by hydropower that had regulated minimum discharges were considered the comparator (because they were most similar to an unimpacted system) while river systems altered by hydropower that had no regulated minimum discharges were considered the intervention. This enabled us to include *ALT-CI* studies during quantitative analysis while ensuring that the direction of expected change was similar to other *CI* studies.

Additionally, all articles included on the basis of full-text assessment underwent quantitative data extraction when possible. No study was excluded from quantitative data extraction based on study validity. Sample size (i.e., number of rivers or sites within a single river) and outcome (reported abundance or biomass metrics) were extracted as presented in tables or text. When studies reported outcomes from multiple sites within comparator or intervention (i.e., different waterbodies or waterbody sections), we averaged these results to obtain a single value. When multiple sampling years (i.e., different after years in a *Before/After* study) or seasons (i.e., sampling seasons within a *CI* study design) were reported separately, we extracted each separately. Data from figures were extracted using the data extraction software WebPlotDigitizer (Rohatgi 2015) when necessary, or authors were contacted to request access to data not otherwise accessible in figures or supplementary figures.

4.6.2 Data extraction considerations

Following full-text screening of articles by the review team, relevant studies and datasets were extracted from included articles. During data extraction, the following considerations defined our database of information. First, we defined a *Site* as a specific hydroelectric facility (i.e., hydro dam) where observations or experiments were conducted and reported in one or more articles (Table 4). Each specific hydroelectric facility was given a "Site ID", using the hydrodam name.

If no name was provided, we used the facility name (if provided) or river name. If no dam, facility or single river name was available, we used NR. Different articles reporting information from the same site were numbered to provide a unique identifier (e.g., Rupert1, Rupert2). Second, we defined a *Study* as an experiment or observation that was undertaken over a specific period at particular sites reported as separate waterbodies that were not treated as replicates within the article. When multiple studies were reported within an article, they were entered as independent lines in the database. Study ID included the Site ID plus a letter (e.g., WreckCove_A, WreckCove_B). Third, a single study could also report separate relevant comparisons, defined as *Datasets* for the same or different species, operating conditions, outcomes, life stages, sampling methods, years/seasons post-treatment and/or sites, but otherwise the same meta-data (Table 4). Each dataset was reported as a separate line in the database and assigned a Dataset ID that included Study ID plus a number (e.g., WreckCove_A1, WreckCove_A2).

If authors reported responses for the same species and the same outcome category in a single study, with otherwise consistent meta-data, we extracted separate datasets for the database when there were different (i) life stages (e.g., the abundance of eggs for species X and the abundance of age-0 for species X), (ii) residency status (i.e., resident vs. non-resident fish); (iii) years and/or seasons post-treatment within a given outcome category (i.e., if for a given outcome category, multiple time periods were monitored and reported separately for a *CI* study design or within-in year variation post-treatment for a *BA* design); (iv) sampling methods (e.g., electrofishing and snorkeling), and/or (v) sites downstream of a hydro dam within a single river sampled using a *BA* design. For quantitative analyses, we aggregated these datasets to reduce non-independence (see “Combining data across multiple comparisons within a study” below).

When a single study reported multiple outcomes within a particular outcome category (i.e., abundance and density or CPUE) but otherwise the same meta-data, for quantitative synthesis we selected the metric closest to the primary focus of the review for each outcome and retained other outcomes only for narrative review. For example, if both abundance and CPUE were reported within a given study, we selected abundance data for quantitative synthesis and retained both abundance and CPUE in the narrative synthesis. Furthermore, if biomass and yield were reported within a single study, we selected only the biomass dataset for quantitative synthesis but retained both biomass and yield in the narrative synthesis (see Appendix 6 for a decision tree). In all cases, for quantitative synthesis we worked to maximize replication, selecting outcomes that had true replication or pseudo-replication over outcomes with no replication. If an outcome metric, closer to the primary focus of the systematic review, had no replication, the next, more 'distant' metric with greater replication was selected by default for quantitative analysis (i.e., if the study reported unreplicated abundance, but density was pseudorelicated, density was selected for quantitative analysis). We retained both outcomes for narrative synthesis in this case as well.

While we extracted all possible data from each study, in some cases it was not possible to retain all datasets extracted. When retaining a dataset would lead to double counting of the same individuals, the aggregated datasets were removed. For example, individual species outcomes were selected over grouped species outcomes for narrative and quantitative synthesis (i.e., abundance of *Oncorhynchus mykiss* were selected rather than *Oncorhynchus* spp. abundance) while still maximizing replication. Similarly, individually reported life stages (e.g., adults) were selected over grouped life stages (e.g., mixed life stages), individually reported monthly data were retained over grouped monthly data, and data for individual sites were retained over aggregated site data. The only exception to this occurred if data for multiple months or sites were

aggregated and data for a single month or site were also reported. In those instances, replication was maximized and individual site data were not considered for either narrative or quantitative synthesis. Two additional reasons for not retaining datasets for either narrative or quantitative synthesis include: (i) no fish of a specific species were captured at a particular site both *Before* and *After* an intervention occurred, or (ii) when samples were taken in a diversion reach (i.e., area downstream of where water is removed from the system) and an outflow reach (i.e., area downstream of where water is returned to the system) and no *Before* data was available for either the comparator or intervention reaches. To ensure independence of the datasets, and based on stakeholder input, only the diversion reach was retained for quantitative and narrative synthesis.

Two types of replication within studies (i.e., group sample size) were considered separately to make use of as much data during quantitative synthesis as possible. Spatial replication in *Control*/*Impact* studies was considered at two levels: (i) independent intervention areas (i.e., separate waterbodies receiving treatments – true replicates) and (ii) subsampled data within rivers, referred to as pseudoreplicates (e.g., multiple samples made upstream and downstream of a dam). Pseudoreplicates have reported variances for the variability among subsamples within a true replicate, rather than the variability among true replicates. For true-replicates, we recorded the number of independent intervention and comparator rivers at the level of true replication, while for pseudoreplicates, we recorded the number of pseudoreplicated samples at the plot or subsample level within the intervention and comparator areas of a single river (i.e., non-independent replicates). We accounted for pseudoreplicated data by making appropriate adjustments during quantitative synthesis (see "Adjustment accounting for pseudoreplication"; Appendix 7). Temporal replication (i.e., for *Before*/*After* intervention study designs) was treated separately (see "*BA* data extraction" immediately below). Temporal replication was considered

here since no *BA* studies included spatial replication (i.e., used a *BA* study design with >1 replicate waterbodies). If only spatial replication was considered for quantitative synthesis, all *BA* studies would have been ineligible for meta-analysis due to lack of replication. Because of this difference in replication between *CI* and *BA* study designs, separate quantitative analyses were conducted for each type of replication (see "Quantitative synthesis").

4.6.3 BA data extraction considerations

We decided whether outcome data sampled during a certain calendar year represented *Before* or *After* as follows:

The *Before* period was defined to date back as long as fish outcome data were available. The *Before* period was defined to end with (and include) the last pre-intervention year. Periods without fish outcome data were included in the *Before* period if they lasted no more than five years and were preceded by a year with outcome data. Furthermore, if there was only a single *Before* period of fish outcome data, as long as it occurred within five years of the start of the intervention, the data were included as the *Before* period.

The *After* period was defined to begin with the first post-intervention year and last as long as fish outcome data were available, and no additional changes/modifications to flow magnitude began. If additional changes/modification were made after the initial period, these *After* periods were also retained and compared to both the *Before* period, and the previous *After* period during quantitative analysis. Periods without outcome data were included in the *After* period if they lasted no more than five years and were followed by a year with outcome data.

In cases where a gap greater than five years occurred, data were extracted, but considered as a deficient *Before/After* comparisons during critical appraisal, and the effect of retaining these

data was explored using sensitivity analysis during quantitative analysis. Gaps longer than five years occurred in four studies (two studies with a *Before* period gap, and two studies with an *After* period gap).

Temporal replication was considered at two levels: (i) within-year (n = # months), and (ii) interannual (n = # years). For within year variation, each *After* year was extracted as a separate row (i.e., different datasets from the same study), with the mean fish outcome and variation for each *After* time period coming from within-year sampling (e.g., averaged across sampling months or seasons). If fish were sampled for only one *Before* year (but for >1 month or season), that *Before* within-year mean and variation were used as the comparator for each separate *After* year. If there were multiple within-year time periods (i.e., >1 year and each year fish were sampled in >1 month), we used the most recent *Before* time period (within-year mean and variation) and recorded this for each separate *After* period. We accounted for multiple comparisons to the same *Before* year during quantitative analysis.

When fish outcome data were available for more than one year in a *Before*/*After* design, interannual replication and calculation of interannual variation allowed us to include these data in separate analyses even if no usable information was available on within-year variation (i.e., when a single fish sampling period occurred per year over >1 years, or when only a total fish abundance for multiple within-year sampling periods was reported for >1 years). Treating within-year and interannual variation separately ensured we did not introduce bias by considering only interannual variation, if within-year and interannual variation differed. For example, if it was suspected that within-year variation in fish abundance was larger than the interannual variation, using effect sizes with interannual variation only would lead to a lower variance and would be given a higher weight in meta-analysis than if the within-year variation

had been known and included as well. Calculations of interannual variation followed two scenarios. First, if (a) fish outcomes were only sampled once per year, or (b) studies only report total fish abundance from multiple sampling seasons within a given year, then mean fish abundance and variation were calculated by averaging these data across all *Before* years (n = # *Before* years), and all *After* years (n= # *After* years). Second, if fish abundance was sampled/reported more than once per year, average abundance was calculated per year (or used in the case where authors reported this average), then averaged across all *Before* years (n = # *Before* years) and all *After* years (n=# *After* years). In the latter case, we were able to make use of studies that reported average fish abundance (from multiple within-year samples) but did not provide any information on within-year variation which would have precluded inclusion in the within-year variation analysis above.

If *BA* study designs were carried out at multiple waterbodies and these waterbodies were not treated as replicates within a given article (i.e., fish responses to changes in magnitude were reported separately for each waterbody rather than combined in a single analysis as replicates), each waterbody was treated as a separate study (same article ID, different Study ID). If a single waterbody was sampled in a *BA* design but at multiple locations downstream of the hydropower dam, each sampling location was extracted on separate rows and treated as the same study [same Study ID, different Dataset ID - similar to different years/seasons sampled in *CI* designs; these were later aggregated to reduce non-independence for quantitative synthesis (see "Combining data across multiple comparisons within a study" below)]. One exception was when both a diversion and a return section were considered in the same study. If this occurred (and a comparator site was available), data from all sites were extracted on separate rows and treated as the same study (same Study ID, different Dataset ID), but the diversion and return sections were

not combined using a composite effect size. If studies combined multiple sites within a given waterbody (i.e., total or average fish abundance), all data were extracted and noted in a comment, then a single metric, maximizing replication was selected for quantitative analysis (see previous section "Data extraction considerations").

4.6.4 Data extraction consistency checking

As described previously (see "Study validity assessment"), to ensure meta-data coding, quantitative data extraction and study validity assessments were extracted in a consistent manner two reviewers (MH and TR) piloted the extraction form by coding and assessing information from 5/103 of the same articles (5%) at the beginning of the process. An additional three articles were used to further test quantitative data extraction. Any disagreements (i.e., what constitutes a high, low or very low head dam; see Appendix 5 for definitions) were discussed and additional detailed guidance was added to the extraction codebook to improve clarity. Coding proceeded with one reviewer (MH) and any queries were discussed with a second reviewer (TR) and a consensus decision made. If a decision could not be reached by the two reviewers (MH and TR) uncertainties were discussed and reconciled with the broader research team and refinements to the coding were made in the extraction codebook as required. Reviewers did not extract data from any study on which they were an author.

4.7 Potential effect modifiers and reasons for heterogeneity

For all articles included on the basis of full-text assessment, we recorded information on the following key sources of potential heterogeneity, if available:

- waterbody type (e.g., river, estuary or canals and diversion channels),
- dam size (i.e., high, low or very low head),

- hydropower operational regime (i.e., run-of-river/modified run-of-river, storage or peaking),
- direction of flow magnitude alteration (i.e., increases/decreases in average, peak, base flow magnitude, increases/decreases in short-term variation, and any combination of these changes in flow magnitude; see Table 4 for types of changes and their definitions),
- alterations to other flow components (i.e., frequency, duration, timing, rate of change, or surrogates of flow alteration, or any combinations of alterations),
- sampling methods [i.e., active or passive gear (electrofishing, net samples, trapping), angling, telemetry, mark-recapture, visual, passive integrated transponders (PIT tags) or others],
- sampling seasons,
- type of comparator [temporal or spatial (upstream of dam, no hydropower - separate but similar waterbodies without HPP, or alternative hydro - separate but similar waterbodies with a different HPP regime)],
- time since intervention (years),
- monitoring duration (years), and
- life stage [i.e., egg: eggs, nests and redds; larvae: larvae, alevins, free embryos; age-0: fry, parr (0+), age-0+, YOY; juveniles: age-1+, parr (1+), juvenile, fingerling (if specific developmental stage is not identified), smolt; adult: adult, spawner, kelt; mixed: assorted life stages].

Potential effect modifiers were selected in consultation with the Advisory Team. When sufficient data were reported and sample size allowed, these potential modifiers were used in meta-analysis

(see the "Quantitative Synthesis" section below) to account for differences among datasets via subgroup analyses or meta-regression (see Table 3 for definitions of terms such as datasets).

4.8 Data synthesis and presentation

4.8.1 Descriptive statistics and narrative synthesis

All relevant studies included on the basis of full-text assessments were included in a database providing meta-data on each study. All meta-data were recorded in a MS-Excel database (Appendix 5) and were used to develop descriptive statistics and narrative synthesis of the evidence, including figures and tables. No studies were excluded from narrative synthesis based on study validity.

4.8.2 Quantitative synthesis

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: (i) lacked replication in the intervention and/or comparator group (i.e., either spatial or temporal); and (ii) did not report measures of outcome variability (i.e., for calculated medians) and/or data on sample sizes and these data could not be otherwise calculated. When possible, imputation (i.e., replacing missing data with calculated substitute values) was used to calculate missing variances (see Appendix 7). Additionally, when only presence/absence data was available for a study or dataset, the study could not be used for quantitative analysis but was retained for the narrative review.

Initial Data Preparation

Outcomes from *BACI* studies were converted to *CI* or *BA* prior to quantitative synthesis to permit analysis with the selected effect size calculations. To convert a *BACI* to a *CI*, data sampled before the intervention (*B*) was subtracted from data sampled after the intervention (*A*) for each

comparator (*C*) and intervention (*I*) site [i.e., *C*: *A-B* and *I*: *A-B*]. To convert to a *BA*, data sampled in the comparator site (*C*) were subtracted from data sampled in the intervention site for the *B* and *A* years [i.e., *B*: *I-C* and *A*: *I-C*]. Means and variances were obtained by averaging across sites or years in each group (see calculations in Appendix 7). In all cases, we worked to maximize the resulting sample size during conversion. Measures of variability were converted to standard deviations when not reported as such (e.g., standard error or confidence intervals) using RevMan Calculator (Drahota and Bellor 2008), and if no variance was reported for averages, standard deviations were obtained using mean value imputation (see calculations in Appendix 7).

Quantitative synthesis—data preparation

Combining data across multiple comparisons within a study

To reduce multiple effect size estimates from the same study and avoid giving studies with multiple estimates more weight in analyses, datasets were aggregated (see Appendix 7 for full description) in five instances when studies sharing all other meta-data, reported: (i) responses from multiple life stages separately within the same outcome (e.g., the abundance of eggs for species X and the abundance of age-0 for species X, separately) (seven studies); (ii) years and/or seasons post-treatment within a given outcome category (i.e., if for a given outcome category, multiple seasons were monitored and reported separately for a *CI* study design (one study; not included for meta-analysis due to lack of replication) or within-year variation post-treatment for a *BA* design) (four studies); (iii) different sampling methods (no studies), and (iv) sites downstream of a hydro dam within a single river sampled using a *BA* design (13 studies), and (v) for one study in which data for both resident and non-resident individuals of the same species were reported and aggregated.

With respect to situation (i) above, most cases were from sampling that was conducted at different time periods within a calendar year (e.g., the abundance of eggs were collected in spring and the abundance of YOY later in summer), potentially resulting in the same individuals being sampled at different time periods and leading to a lack of independence between responses. If sampling was conducted at the same time period, responses could have been summed using equations 23.1-23.3 in Borenstein et al. (2009a); in the one *CI* instance where this occurred it was not possible to sum the responses of the different life stages for three reasons: (1) this study was a *BACI* conversion to *CI*; (2) outcomes were reported as densities; (3) intervention sites were pseudo-replicates. If results had been summed, this would have led to an inflated sample size and required an unacceptable level of data manipulation. Therefore, we aggregated such cases instead.

With respect to situation (ii) (i.e., aggregating by years and/or seasons post-treatment within a given outcome category) we attempted to aggregate across seasons, but only a single study with insufficient replication reported seasonal data requiring aggregation. If data were reported for multiple post-treatment years for a *CI* and magnitude remained status quo (meaning no additional changes occurred since time period/year-1), we only included the first year post treatment (year-1) in the main analysis but extracted separate datasets for all cases where there were more years of data, in order to explore the possibility of a temporal lag in fish responses to flow magnitude alterations.

In some cases, more than one situation for aggregating datasets was present (i.e., multiple life stages and downstream sites were reported separately). When this occurred, we did not average across both life stages and downstream sites, because outcomes were already averaged across years, leading to unacceptable levels of data manipulation. To avoid this, we selected the

life stage that was less likely to be under-represented by the sampling technique used [i.e., catchability by electrofishing increases with size (Borgstroem and Skaala 1993; Hedger et al. 2018)]. Similarly, if a single species was sampled in one location, but different life stages were captured using different techniques, we aggregated by life stage, disregarding the sampling technique, under the assumption that the authors selected the sampling technique that was most effective for each life stage targeted.

Handling dependence from multiple group comparisons

In our database of effect sizes, there were a few instances of multiple group comparisons whereby related studies used a single group of control sites and more than one operational regime [e.g., peaking, run-of-the river, storage (two studies each)] or had multiple interventions comparable to both the initial *Before* period and to previous *After* periods (two studies). In some cases, a single comparator site upstream of a diversion and return reach was compared to these two intervention types (four studies). In such cases, the comparator group was used to compute more than one effect size and, in consequence, the estimates of these effect sizes are correlated. Because we were interested in testing for the association between operational regime and effect size, we did not aggregate multiple group comparisons within a single study across these regimes. To reduce such case dependencies, we would have removed datasets for a given operational regime where there were insufficient combinable data (i.e., < 3 datasets from < 2 sites); however, this did not occur so no datasets were removed. For cases of multiple group comparisons, we performed sensitivity analyses to compare models fitted with and without such cases to examine differences in pooled effect sizes.

Effect size calculation

Because outcomes (e.g., abundance, density, CPUE, or biomass, yield) were not always reported in comparable units or on the same scale, we used the standardized mean difference, Hedges' g (Hedges 1981), as our effect size measure rather than raw mean differences. Hedges' g was calculated using the steps in Borenstein et al. (2009a), as shown below.

Starting with Cohen's d to take account of differences in measurements across studies (Cohen 1977), we calculated the standardized mean difference by dividing the mean difference in each study [i.e., the difference between mean fish responses to an intervention and the mean fish response to a lack of intervention (the comparator)] by the study's pooled standard deviation:

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

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

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

where S = standard deviation, and n is the sample size. The variance for d is given by:

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

Then, to covert Cohen's d to Hedges' g, we used a correction factor that decreases small sample bias in d:

$$J = 1 - \frac{3}{4(n_{G1} + n_{G2} - 2) - 1} \quad (4)$$

Finally, we calculated Hedges' g and the associated variance (V_g) as:

$$Hedges'g = J \times d \quad (5)$$

$$V_g = J^2 \times V_d \quad (6)$$

From this, a negative Hedges' g indicates that fish outcomes (abundance or biomass) are lower after the intervention (i.e., sites impacted by a change in flow magnitude or the *After* period of a *Before/After* study), than in the associated comparator (i.e., sites unimpacted by a change in flow magnitude or the *Before* period of a *Before/After* study). Adjustments were made to these equations when conducting calculations for studies with pseudoreplicates (see "Adjustments accounting for pseudoreplication" in Appendix 7).

<u>*Quantitative synthesis*</u>

To determine if changes in flow magnitude had an effect, on average, on fish abundance and biomass outcome metrics, fish responses were compared to controls by conducting random-effects meta-analyses using restricted maximum-likelihood (REML) to compute weighted summary effect sizes for each outcome (i.e., abundance and biomass) within a given replication type (i.e., spatial replication for *CI* study designs, within-year and interannual temporal replication for *BA* study designs) separately. A random-effect model assumes that there is no true effect size that is fixed for all studies and instead assumes that effect sizes will be different, but similar, across studies and that the effect sizes are a random sample from a population of effect sizes (Hedges and Vevea 1998; Borenstein et al. 2010).

For within-year *BA* comparisons, models were developed for each of the first four years after a change in flow magnitude [i.e., comparing the most recent or only *Before* year with (i)

After year-1 only, (ii) *After* year-2 only, (iii) *After* year-3 only, and (iv) *After* year-4 only], as well as the average of years 1-4 after a change in flow magnitude (see Additional File 7 for full description). The first four years after a change in magnitude were selected since there were insufficient sample sizes in the available evidence base beyond this time frame.

To account for species outcomes reported from the same site but from different studies (see Appendix 7 for full adjustment summary), Study ID was included as a random factor in each model. The summary effect size was considered significantly different from zero when the 95% confidence interval (CI) did not overlap with zero. Heterogeneity in effect sizes was calculated using the Q statistic, compared to the χ^2 distribution to determine if the total variation in observed effect sizes (Q_T) was more heterogenous than expected due to sampling error alone (Q_E) (e.g., Q_T is significantly greater than expected from Q_E) (Rosenberg 2013). A statistically significant Q indicates greater heterogeneity in effect sizes (i.e., individual effect sizes do not estimate a common population mean), which suggests there are differences among effect sizes that arise from causes other than sampling error. We produced forest plots to visualize mean effect sizes and 95% CI from each comparison. Summary effect sizes were used to identify general trends in the evidence base and the impact of the intervention. 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.

Although we attempted to reduce publication bias by including data from available grey literature, publication bias could still impact results if publishing is biased towards a particular type of result, such as statistically significant outcomes. Therefore, we examined publication bias for global analysis models (as described above) by testing for this bias using funnel plots and

fail-safe numbers. Visual assessment of funnel plots (scatter plots of included studies' effect sizes versus a measure of precision such as sample size, standard error or sampling variance; Light and Pillemer 1984) was used to determine if bias was present. If no bias was present, the funnel plot should be funnel-shaped, with wider spread of effect sizes with lower precision (i.e., smaller studies) and less spread as precision increases (i.e., larger studies) (Light and Pillemer 1984). We used funnel plots where precision was based on 1/square root of sample size (*k*), because funnel plots based on sample size are less susceptible to distortion than those based on standard error (Zwetsloot et al. 2017). In these plots, as sample size increases and $1/\sqrt{k}$ decreases, the variance in the effect sizes is expected to decrease if publication bias is not present. In addition to funnel plots, we used fail-safe numbers to test the robustness of our results against publication bias using the method described in Rosenberg (2005), and the *fsn* function in the metafor R package (Viechtbauer 2010). Fail-safe numbers indicate the number of nonsignificant, unpublished (or missing) studies needed to eliminate a significant overall effect size (Rosenthal 1979; Rosenberg 2005). The failsafe number is considered robust if it is greater than $5k + 10$, where k is the number of effect sizes in the analysis (i.e., it is unlikely that the number of unretrieved studies is five times that which are considered in the review and the minimum number likely missed is set at 10; Rosenthal 1991).

To test for associations between effect size and moderators, we used mixed-effects models for categorical moderators [i.e., (i) waterbody type, (ii) dam size, (iii) hydropower operational regime, (iv) direction of flow magnitude alteration, (v) alterations to other flow components, (vi) sampling methods, (vii) sampling season, (viii) type of comparator (temporal/spatial), (ix) study class (manipulative vs. nonmanipulative), (x) time since intervention, (xi) monitoring duration (*CI* only), and (xii) life stage] and meta-regression for

continuous moderators (i.e., monitoring duration; *BA* only), when possible. We estimated heterogeneity in these models using REML. We only performed analyses for categorical moderators when there were sufficient combinable datasets (i.e., ≥3 datasets from at least 2 studies) for each moderator category (e.g., at least 3 datasets from at least 2 studies for each of the operational regimes *Peaking* and *Storage*). In some cases, there were insufficient numbers of datasets in different moderator categories; therefore, categories were either combined with similar categories to increase sample size [e.g., when there were insufficient datasets with the same alterations to flow magnitude elements (i.e., Baseflow or AvgDischarge), flow elements were combined into larger groups based on the direction of alteration (increase or decrease)] or datasets were deleted if they did not meet sample size criteria (see details in "Results"). For example, studies with comparator sites in waterbodies with alternative levels of hydropower could not be combined with studies where comparator sites were in waterbodies without hydropower; therefore, these datasets were deleted from analyses.

Because studies did not always report all moderators of interest, it was not possible to combine all moderators into a single model simultaneously, nor did sample size allow this. Therefore, to test associations between effect size and moderators, we first conducted random-effects models (unmoderated models) using subsets of responses (e.g., a subset of abundance or biomass effect sizes for a given replication type) that maximized the number of effect sizes that could be used to test the influence of the moderator of interest. We then used these subsets in mixed-effects models or meta-regression, including the moderator of interest. To further account for multiple study comparisons within a study site, and species outcomes being reported for the same site, in all models, Study ID was included as a random variable. We restricted the number of fitted parameters (j) in any mixed model such that k/j where k is the number of effect sizes,

was greater than five to ensure reasonable model stability and sufficient precision of coefficients (Vittinghoff et al. 2005). This limited the number of moderators and categories that could be included in a single model. Given that all moderators were highly correlated (see results of Pearson's χ^2 test of moderators; Appendix 12: Table S1 and S2), it was not possible to add more than one moderator into a given model, nor would sample size allow for this.

For all moderator analyses, total heterogeneity (Q_T) was partitioned into the heterogeneity explained by the model (Q_M) and heterogeneity not explained by the model (Q_E, error due to sampling); therefore, $Q_T = Q_m + Q_E$. The statistical significance of Q_M and Q_E were tested against a χ^2 distribution. For *CI* studies, monitoring duration was treated as a categorical variable because of low variability in studies of short duration and few representative longer-term studies (i.e., one study at five years and one study at 36 years duration). Because of two outliers, it was not possible to transform monitoring duration for *BA* studies to meet model assumptions and reduce skewness, while maintaining all datasets; therefore, we conducted mixed-effects models including monitoring duration with and without outliers. Results did not differ (see Appendix 11) and we only present results without outliers below.

Sensitivity analyses

Sensitivity analyses were carried out to investigate the influence of: (i) study validity categories; (ii) imputing missing variances; (iii) inclusion of studies where the waterbodies may be influenced by fish stocking; (iv) inclusion of studies with pseudoreplication (*CI* studies only); (v) inclusion of multiple group comparisons where a single comparator group was compared to more than one intervention group within the same study and outcome/type of replication subgroup; (vi) inclusion of articles that did not specify a flow magnitude component or reported unspecified multiple components of flow (*CI* studies only); (vi) inclusion of deficient *BA* or *BACI* study

designs (*BA* studies only); (vii) inclusion of yearly averages, averaged for the *Before* or *After* period (i.e., averages of averages; *BA* studies only), and (viii) inclusion of outfall zones (*BA* studies only). First, models were fit with only those studies assessed as 'Medium' validity. Second, separate models were fit using only studies with variances that did not require imputation during data preparation. Third, separate models were fit using only studies where stocking was not known to be a potential confounder (i.e., we did not include studies where authors indicated that stocking may have occurred in the system but did not include sufficient information to determine if stocking was ongoing). Fourth, separate models were fit using only studies with true replication for studies with spatial replication (*CI* studies only). Fifth, separate models were fit with only studies with a single intervention and a single comparator. Sixth, we ran separate models with studies that specified flow magnitude (i.e., we did not include studies that compared an unregulated stream or stream section to a regulated stream). Seventh, we ran separate models (for *BA* studies only) that did not use *DEF_BA* designs. Finally, we ran separate models (for *BA* studies only) that did not include both an abstraction and an outflow reach. In all analyses, the results were compared to the overall model fit to examine differences in pooled effect sizes. All meta-analyses were conducted in R 4.0.3 (R Core Team 2020) using the *rma.mv* function in the metafor package (Viechtbauer 2010).

5. Review findings

5.1 Review descriptive statistics

5.1.1 Literature searches and screening

Updated searches in six databases and Google Scholar resulted in 2966 individual records (Fig. 2). Other websites and databases suggested by experts and practitioners identified 1695 individual records. From the 2966 records identified from the six databases and Google Scholar,

721 duplicates were removed using EPPI Reviewer prior to title and abstract screening. No duplicates were identified from the three websites/databases suggested by experts and practitioners. A total of 3940 articles moved forward to title and abstract screening, at which time an additional 21 duplicates were manually identified and removed.

Title and abstract screening removed an additional 2864 articles, leaving 1055 for full text screening; 152 from the six databases and Google Scholar, and 903 from the other websites/databases listed above. Of these articles, 17 were not obtainable because of insufficient bibliographical information or articles were not accessible with Carleton University's subscriptions, leaving a total of 1038 articles for full text screening. An additional 95 articles from pre-screened sources were included at this stage from searching the bibliographies of (a) relevant articles identified (24 articles), and (b) the 110 relevant reviews found with searches (nine articles), and grey literature sources and submissions obtained via social media/email (62 articles).

Full-text screening removed 922 additional articles (878 articles and 44 previously missed duplicates). The majority of articles excluded at full-text were excluded because of an irrelevant study design (i.e., spatial or temporal trends), irrelevant comparator (i.e., downstream or lacking comparators) or an irrelevant intervention (i.e., dams without hydro) and were primarily from grey literature sources (763/878 articles; 86%) (see section "*Eligibility Criteria*" for inclusion/exclusion requirements). A total of 211 articles were included from full-text screening; most were from websites/databases suggested by experts (157 articles), followed by those identified in reference lists of included articles (24 articles), from database and the search engine (21 articles) and submissions from experts (nine articles). All articles excluded at full text, and those that were unobtainable, are listed with an exclusion decision in Appendix 4. An

additional 107 articles identified from the Rytwinski et al. (2020b) systematic map were included for data extraction and were screened for inclusion at the data extraction stage.

A total of 318 articles were initially included for data extraction. During data extraction, a total of 146 additional articles were excluded, including 67 articles that were supplementary to another excluded article (Fig. 2). Although including diversity metrics was our original intent for this review (see published protocol; Harper et al. 2020), due to resource and time constraints, the narrative and quantitative syntheses consider only studies focused on abundance and biomass metrics. This resulted in a total of 103 articles with 134 studies included in the narrative analysis for abundance and biomass. One *CI* study was later excluded prior to narrative and quantitative analysis because the article considered both a diversion and an outflow reach, the effects of which could not be separated, resulting in impacted reaches that were not independent (i.e., study considering outflow reach was removed to ensure independence of studies). Of the remaining 103 articles and 133 studies, all were used in the narrative synthesis and 46 articles with 58 studies were included in quantitative synthesis.

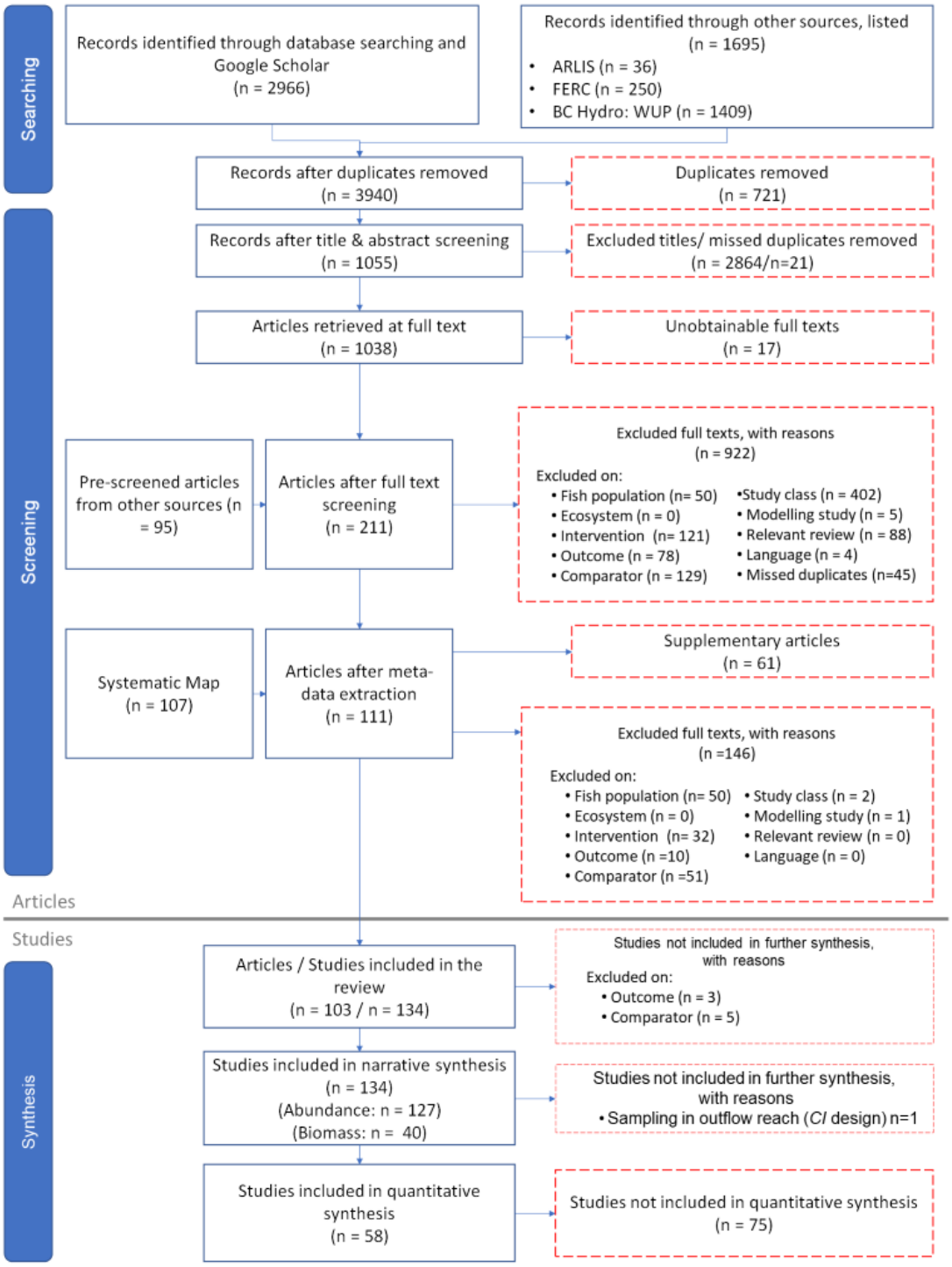

Fig. 2. ROSES flow diagram (Haddaway et al. 2018) showing results of the literature search and study selection process showing the final number of studies included in the systematic review. Blue indicates articles/studies proceeded to next stage of review, red dashed lines indicate articles/studies were removed from consideration at that stage.

5.1.2 Study validity assessment

Validity assessments (see Table 2 for definitions) of the 133 studies resulted in 185 individual projects (Appendix 8). Most projects were assigned an overall 'Low' study validity (131 projects; 71%), with the remaining projects being assigned an overall 'Medium' study validity (54 projects; 29%). No study was assigned an overall 'High' study validity (see Table 5). For all decades considered, 50% or more of relevant projects had 'Low' validity (Fig. 3).

Among the projects that received an overall 'Low' validity score, most (52%) had confounding factors (i.e., manipulations of other flow regime components) or there was a lack of information to judge whether confounders were present. An additional 34.6% of projects that received a 'Low' validity score lacked replication (either spatial and/or temporal). This included studies that: (i) had no replication (29.2%), or (ii) lacked sufficient information to judge replication (5.4%). Among projects that received a 'Medium' validity score (all other studies), the most common reason (25%) was a lack of true replication (i.e., experimental/observational units were pseudoreplicates) and an additional 12% of projects lacked quantitative measures of flow magnitude alterations. Of the 12 studies (20 projects) that used a *BACI* design, none had a 'High' validity score primarily because of insufficient information about the intervention [i.e., no quantitative measure of flow magnitude (nine projects), or compared unregulated to regulated systems but did not specify a change in flow magnitude (10 projects)]. Other reasons for 'Low' and 'Medium' scores for *BACI* studies were distributed relatively equally in all validity categories. Three *BACI* projects from two *BACI* studies, while having sufficient replication in the intervention zone, lacked sufficient information on interventions and confounders to receive a high validity score. These two also lacked sufficient data for the control site before the

intervention occurred (i.e., single year of data) to be treated as *BACI* studies in quantitative analysis.

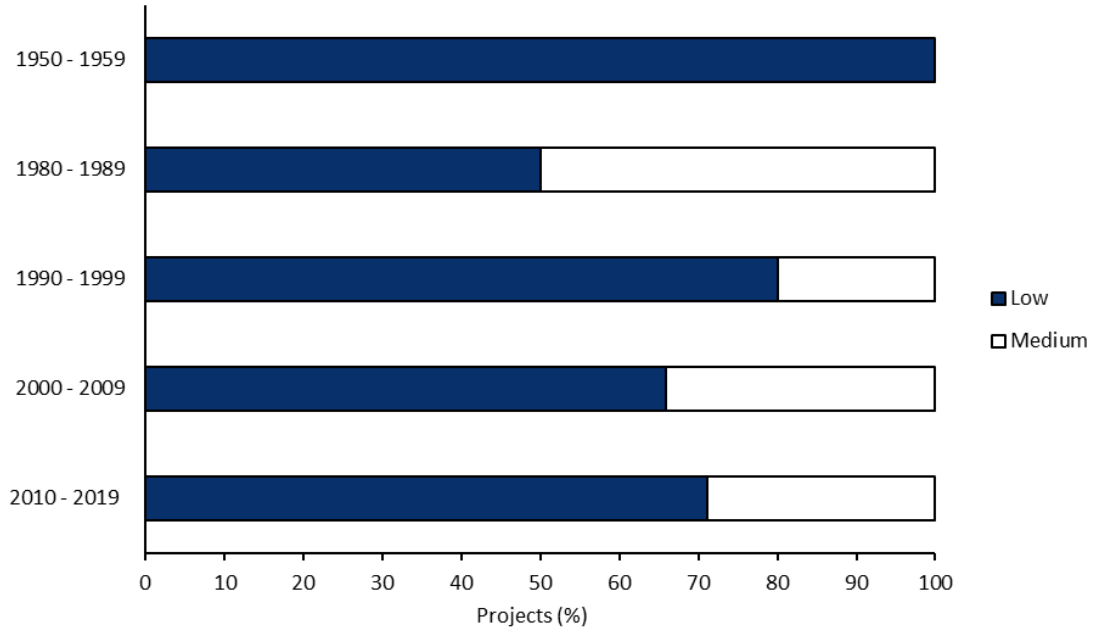

Fig. 3. Study validity of 185 projects in relation to the decade of publication, reported as a percentage of all projects for that decade. No projects from 1960-1979 were included.

Table 5. Results of study validity assessment of the 185 projects considered, using the critical appraisal tool (see Table 2).

Criteria	Response	Projects (#)
Study Design	BACI, RCT	20
	BA, CI, ALT-CI, RCA, NR, INCOM-BACI	156
	DEF-BA or lacking sufficient information to judge	9
Replication at level of intervention	≥2 independent experimental/observational units	99
	At least 2 non-independent experimental/observational units	22
	No/unclear replication in experimental/observational units, or insufficient information to judge	64
Comparator matching*	Well matched at site selection and/or study initiation	27
	Moderately matched at site selection and/or study initiations	26
	Poorly matched at site selection/initiation, or insufficient information to judge	26
Intervention description	Clear description of change and quantitative data	92
	Clear description of change but no or unclear quantitative data	56
	Compares unregulated vs regulated streams (with no reported measure of magnitude) or reports unspecified multiple components affecting flow	37
Confounding factors	No or minimal confounding factors, or factors are accounted for appropriately	88
	N/A	---
	Present and/or not accounted for, or insufficient information to judge	97
Sampling design	Consistent in space and time	127
	Inconsistent through space and time	32
	Lacking sufficient information to judge	26

*Number of projects in this criteria is less than the total number of projects (185) because criteria was not applicable to 106 BA projects. All other criteria sum to 185 projects.

5.1.3 Publication year

Articles included for abundance and biomass metrics were published from 1958 to 2019. Only one article published prior to 1980 was included. The number of publications increased over time, with more than twice the number in the most recent decade compared to any previous decade (Fig. 4). From 2000 to 2019, the quantity of grey literature increased and made up just under half (44%) of all articles.

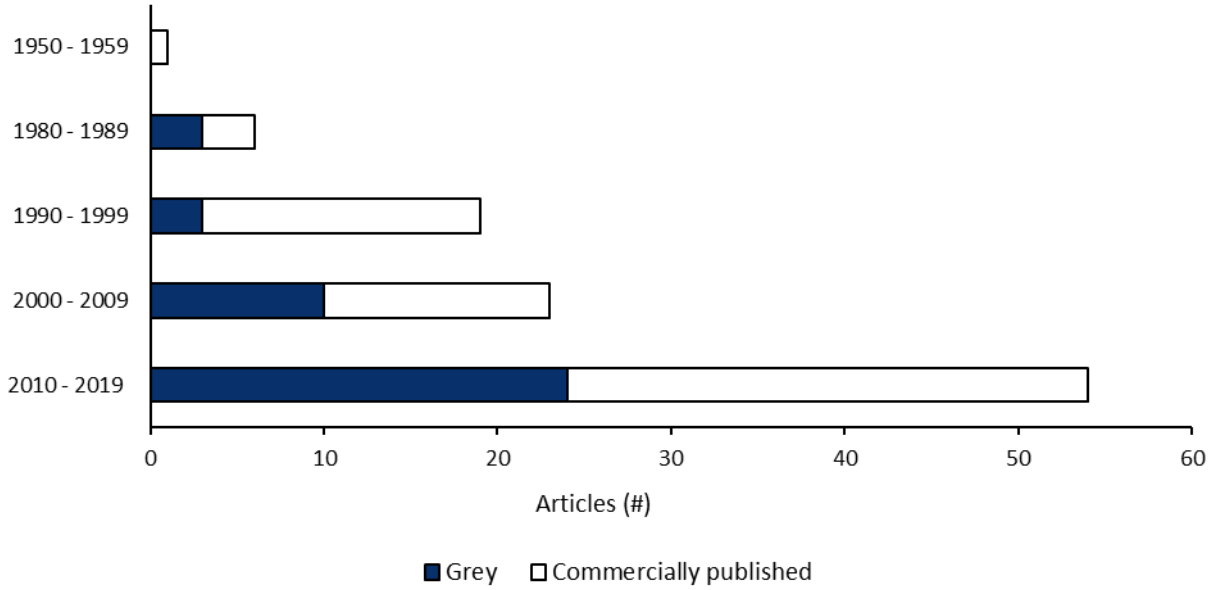

Fig. 4. Frequency of grey and commercially published literature considering abundance and biomass and included for data extraction and critical appraisal in each decade. No articles from 1960 -1979 were included.

5.2 Narrative synthesis

The narrative synthesis is based on all 133 studies from 103 articles that considered abundance and biomass, regardless of study validity. A database of these studies with descriptive meta-data, coding and qualitative/quantitative data is available in Appendix 5.

5.2.1 Study location

Studies occurred in 22 countries (Fig. 5), with most studies conducted in North America (60%); 40 studies were conducted in each of the United States (30%) and Canada (30%). The two most represented American states were California (abundance: nine cases, biomass: five cases), followed by Alabama (abundance: seven cases), and the two most represented Canadian provinces were Quebec (abundance: 21 cases; biomass: three cases) and Ontario (abundance: 16 cases; biomass: 16 cases) (Fig. 6a and 6b). Some studies had sampling sites in more than one state or province, so the number of cases exceeds the total number of studies. Of the remaining 40% of studies, 26% were conducted in Europe (35 studies), 10% were conducted in Asia (13

studies), 2% were conducted in South America (three studies), and Eurasia and Oceania each represented 1% with one study apiece (Fig. 5).

All studies were field based and occurred in river systems. A single study had sampling sites in both river and estuary environments. Studies reported a total of 111 named hydroelectric power dams/facilities. Several studies considered more than one hydroelectric power dam/facility (i.e., one study considered a total of 17 hydropower facilities and two additional unnamed facilities across 16 waterbodies), resulting a total of 146 cases. Several studies did not report the name of any dam or facility (23 studies). The three hydroelectric power dams/facilities considered by the largest number of studies were Rupert Dam in Canada (six studies), Glen Canyon Dam in the United States (five studies) and Three Gorges Dam in China (five studies). A total of eight hydroelectric power dams/facilities were included in three or more studies (Fig. 7), nine were considered in two studies and the remainder occurred in a single study each.

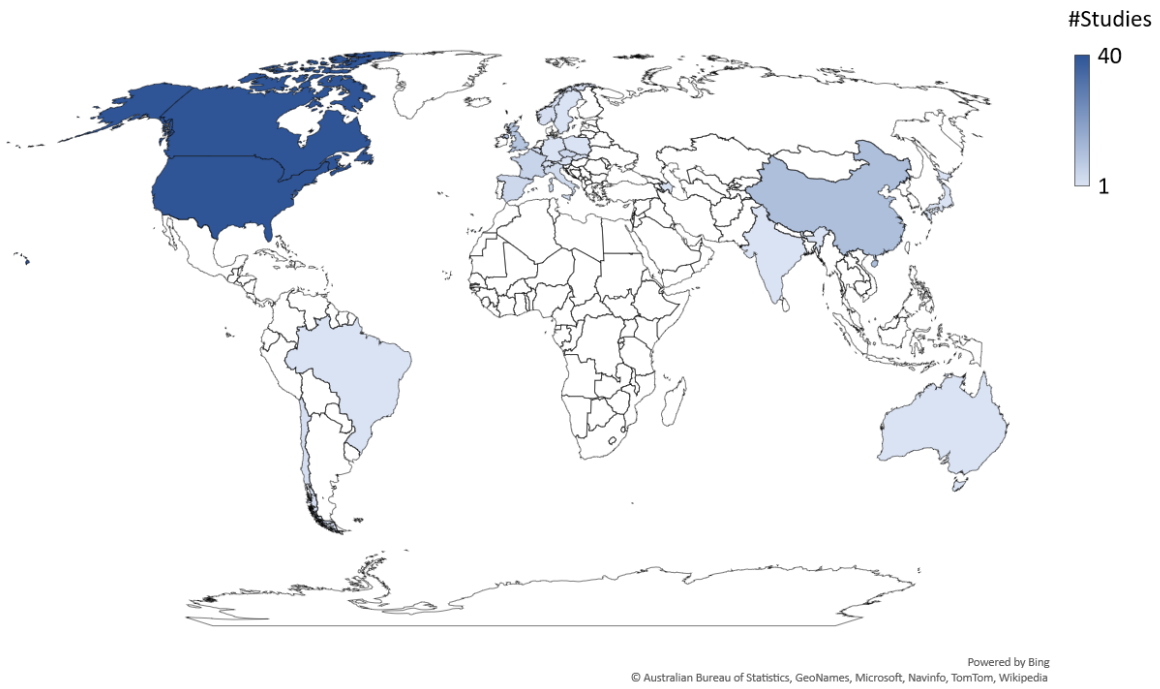

Fig. 5. Number of studies considering fish abundance and/or biomass metrics per country.

Fig. 6. Number of cases considering fish abundance (blue) and biomass (green) metrics per state/province in (a) Canada and (b) the United States. Note the different colour ranges in each map.

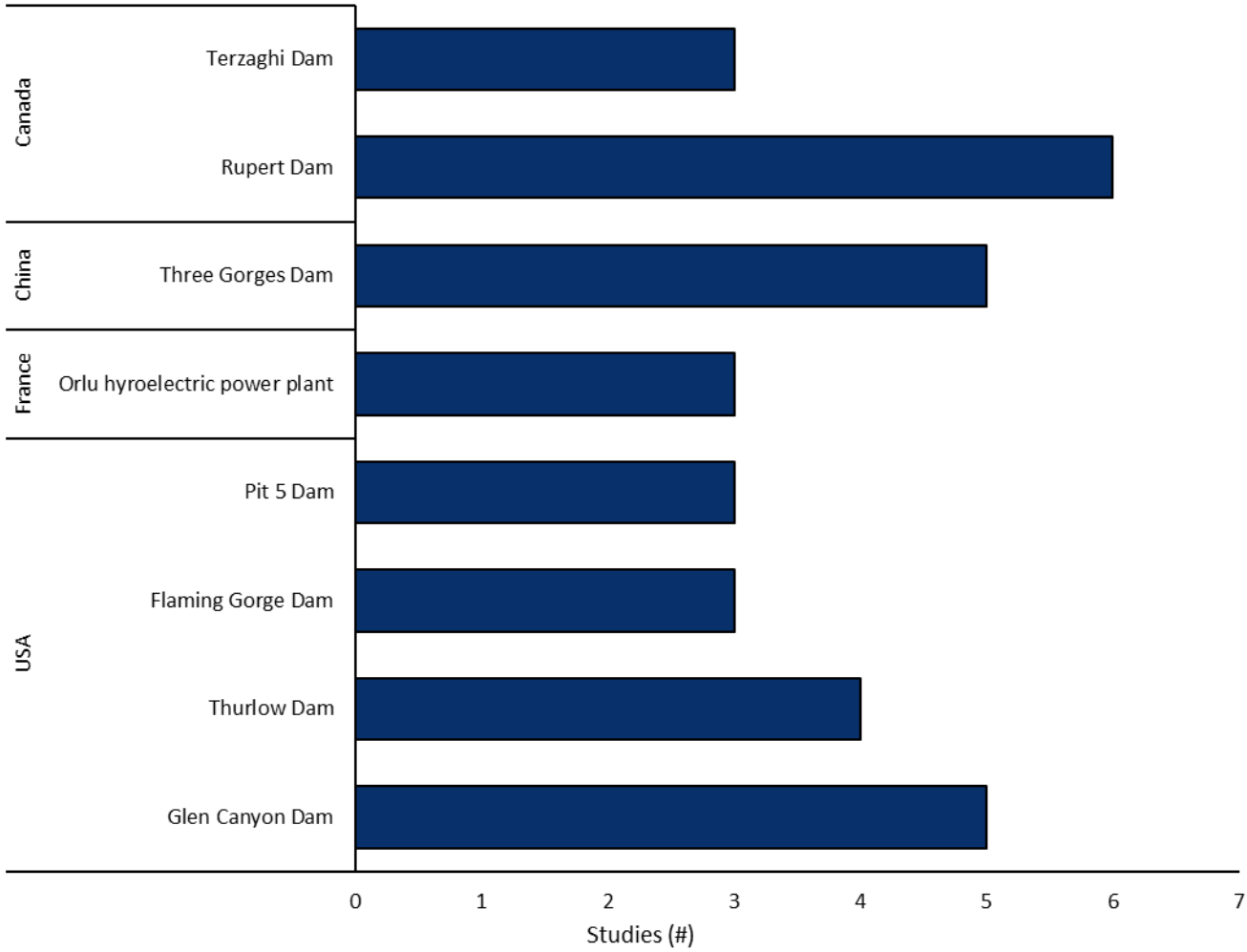

Fig. 7. Eight most studied hydroelectric dams/facilities by country.

5.2.2 Population

Most studies (75%; 100/133) conducted species-specific investigations (i.e., provided data for individual species rather than grouped/pooled over broader categories of species, genus or family). Of studies that reported species specific data, 37% (37/100) considered only one species. The maximum number of species considered by a single study was 85. A total of 47 families were investigated by studies considering the impact of flow magnitude changes on specific species (i.e., not grouped across species, genus or family). This represented 124 genera and 333

species. Studies also reported four unidentified families, 17 unidentified genera from identified families and five unidentified species from identified genera. The top 10 families and their top studied genera are shown in Fig. 8. Salmonidae was the most studied family, of which the most frequently studied genera were *Salmo* (45 studies), *Oncorhynchus* (30 studies) and *Salvelinus* (nine studies). The next most targeted family was Cyprinidae, although all genera were represented by 12 or fewer studies. A total of nine species of Catostomidae were considered (24 studies). Cottidae, while represented by a single genus (*Cottus*), was considered in 23 studies and Percidae were considered in 22 studies. All other families were included in fewer than 20 studies each. The most studied species were *Salmo trutta* (41 studies), *Oncorhynchus mykiss* (24), *Micropterus salmoides* (13), and *Salmo salar* and *Lepomis macrochirus* which were reported in 12 studies each. Of studies that reported species-specific results, two studies did not report enough information to determine which family group a recorded species or species group belonged to (i.e., undefined larval stage). An additional six studies did not report sufficient information to determine the genus either because it was not possible to identify beyond family for the life stage (two studies), or results were for mixed-genus groups. A total of nine studies did not report species names but grouped species under genera. A total of 12 studies reported data that were grouped across family or genus. These studies were not considered in Fig. 8 but did include an additional 30 species not previously reported by articles with species-specific results (Appendix 9). A single study considered an identified hybrid.

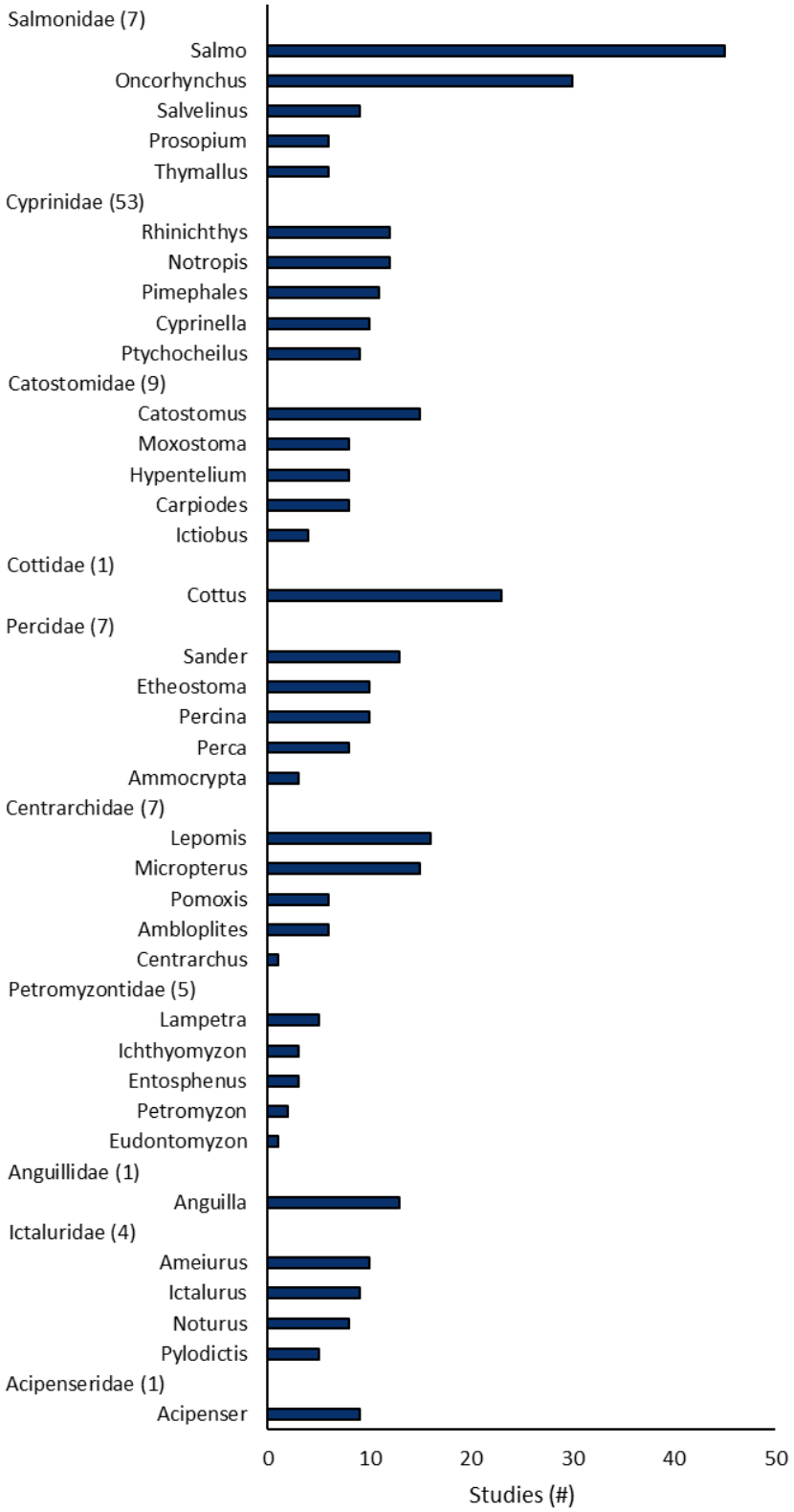

Fig. 8. The number of studies per family and genus of the top 10 most studied families, and their associated top five studied genera. The number of genera per family is shown in brackets adjacent to family name. The number of studies shown exceeds the number of included studies because many studies considered multiple genera.

5.2.3 Intervention

We separated dam type based on head height and operational regime (i.e., peaking, run-of-river, storage). Head height was grouped as: (i) 'high head' dams (>10 m head) (109 cases), (ii) 'low head' (<10 m head) (12 cases) and (iii) 'very low head' dams (<5 m head) based on author descriptions (four cases). In 36 cases, no dam head height was included and there was insufficient information to determine the dam head height from other sources. In terms of operational regime, one study considered more than one operational regime due to the inclusion of a diversion (run-of-river) and outflow reach (peaking), resulting in 134 cases from 133 studies. Operational regimes were relatively equally represented by peaking (38 cases), storage (36 cases) and run-of-river (35 cases) (Fig. 9), although 30 cases did not report an operational regime. When both head height and operational regime were reported, high head dams had a fairly even split of operational regimes for peaking (33 cases), storage (30 cases) and run-of-river operations (25 cases). Low head and very low head dams were primarily storage (five cases) and run-of-river operations (four cases), with peaking regimes represented by only three cases. In 25 cases the operational regime was either not reported or was not clearly described.

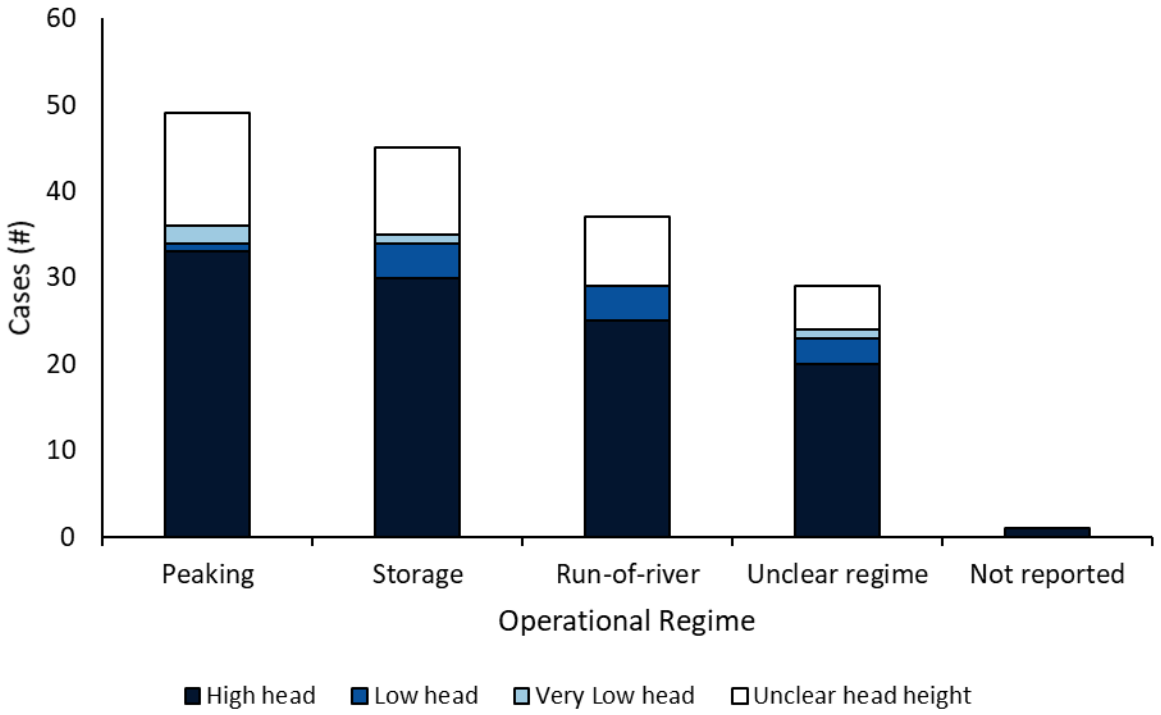

Fig. 9. Number of cases of hydroelectric power production facility operations (peaking, storage, run-of-river) in relation to dam head height: *high* (>10 m); *low* (<10 m); and *very low* (<5 m) head height. *Unclear regime*: type of hydropower production facility operation was not clearly enough described to be classed with other operational regimes; *Not reported*: no information on type of operational regime included.

Of the 133 studies included for narrative synthesis, 12 studies included more than one dam, while 23 studies did not report the name of the facility/dam considered. A total of 70 hydropower facilities/dams were reported independently (i.e., no more than one dam considered within the study) and of these seven dams/facilities were considered by more than two studies (Fig. 10). Some studies included more than one study design (i.e., *BA* and *DEF_BA*) resulting in more cases per dam than the total number of studies. The most studied dams were the Glen Canyon Dam and Rupert Dam (both in North America), with five studies each, however two studies for Rupert Dam reported more than one study design (seven cases).

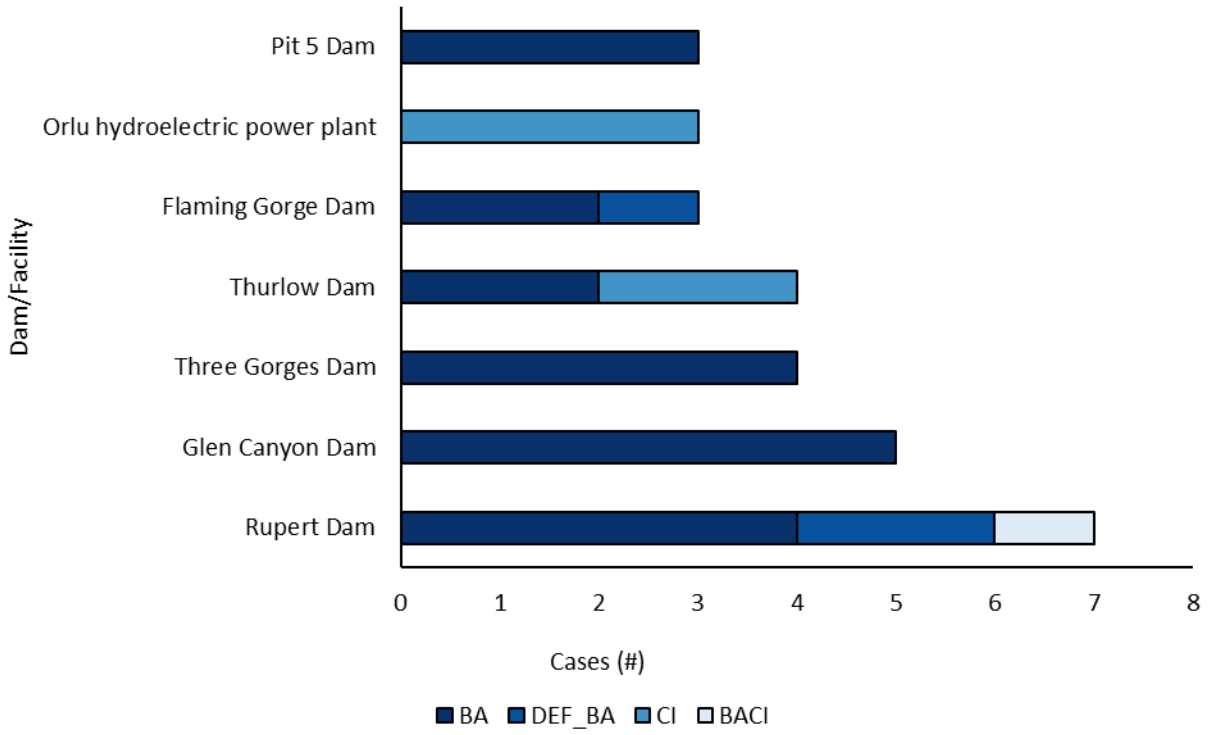

Fig. 10. Dams/facilities considered in more than one study with study design. *BA*: *Before/After*; *BACI*: *Before/After/Control/Impact*; *CI*: *Control/Impact*; and *DEF_BA*: deficient *Before/After* design (i.e., gap of more than five years, or missing information).

Alterations to flow magnitude were generally increases (49 studies), primarily to peak flow (12 studies) or to two or more flow magnitude elements (19 studies) (Fig. 11; see Table 4 for definitions). A total of 43 studies considered decreases in flow magnitude elements, while 26 did not specify any flow element or direction of alteration (Fig. 11). When two flow elements were changed together, normally this included either increasing or decreasing all elements considered (increase: 14 studies, decrease: two studies), although in six studies, one element was increased, while another was decreased (e.g., short-term variation increased, average discharge decreased: two studies). The most common two-element combination was increases to both base flow and peak flow (five studies), or peak flow and short-term variation (five studies). The most common three-element combination was increases to average discharge, peak flow and base flow (four studies) and only a single study had both increases and a decrease in flow magnitude

elements. Two studies had alterations to all four flow magnitude elements (one increased all elements, and the other increased base flow and short-term variation while decreasing average discharge and peak flow).

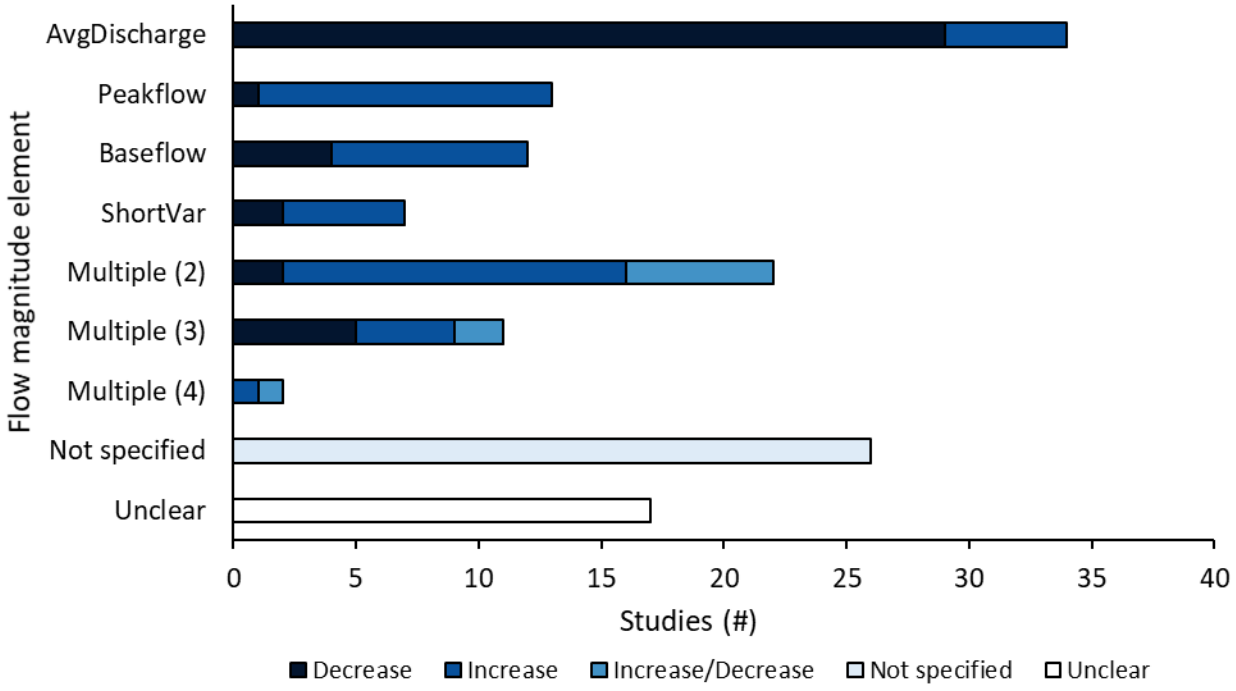

Fig. 11. Number of studies with alterations to the four flow magnitude elements and the direction of alteration (refer to Table 4 for definitions). Changes to flow elements include: average discharge (AvgDischarge), peak flow (Peakflow), base flow (Baseflow) and short-term variation (ShortVar). *Multiple* indicates that more than one flow element was changed. Flow magnitude elements could be increased or decreased. In cases where multiple changes occurred, individual elements could increase and/or decrease separately (i.e., increase/decrease). *Unclear* indicates descriptions were provided by authors but were insufficient, while *Not specified* indicates that no descriptions of flow magnitude element or direction of change were provided.

5.2.4 Study design and comparator

All 133 studies (from 103 articles) were field-based, with no laboratory studies included during full-text screening. A total of three studies reported more than one study design (one study reported both a *BACI* and *BA*, one study reported both *BA* and *DEF_BA* study designs, and one study reported *BACI* and *CI* study designs). Therefore, the number of cases of study designs

exceeds the total number of studies. Most cases had *BA* designs (*BA*: 63 cases; *DEF_BA*: eight cases), followed by *CI* study designs (*CI*: 48 cases, *ALT-CI*: one case) (Fig. 12). Two studies had *RCA* designs and were combined with *CI* study designs during quantitative analysis. *BACI* designs (13 cases) and a single incomplete *BACI* design made up the remainder. No randomized control trials (*RCT*) or normal range (*NR*) studies were included. Some studies originally included from the systematic map were missing required study criteria (e.g., appropriate comparators, interventions, or populations) and were removed during data extraction (40 studies).

The most common temporal comparator used by *BA* or *BACI* studies were periods before the installation of a new HPP facility (42 studies), followed by *Before* periods where altered flow previously existed in the system (41 studies) (Fig. 12). Comparator sites in systems without HPP were most used for *CI* studies (24 cases) and were the only comparator used for *RCA* studies (two cases). *CI* study designs also used upstream comparators (22 cases) (Fig. 12). Additionally, two cases combined upstream/downstream comparators. Only one study used systems with alternative HPP (one *ALT-CI* study) as a comparator. *BACI* and incomplete *BACI* studies used only *Before* periods prior to the installation of a new hydropower system (14 cases), but used both upstream comparators (nine cases) and comparators in systems without HPP facilities (four cases). Although two studies were reported as having *BACI* designs, because of a lack of information about their spatial comparator, they were converted to *BA* designs for quantitative analysis. No *RCT* or *NR* studies were identified.

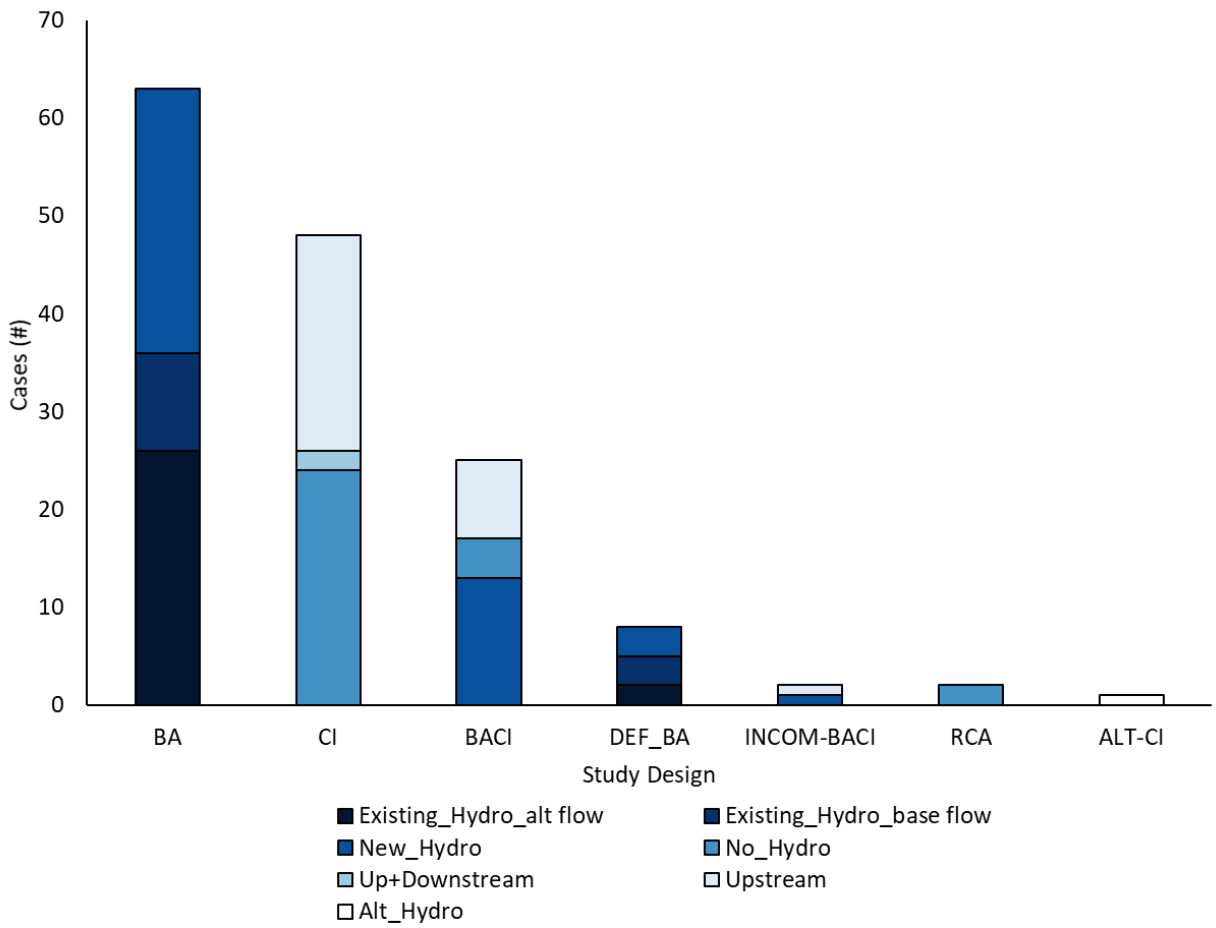

Fig. 12. Number of cases by study design and comparator. Study design codes: BA: Before/After; CI: Control/Impact; BACI: Before/After/Control/Impact; DEF_BA: deficient Before/After, INCOM-BACI: incomplete Before/After/Control/Impact; RCA: Reference Conditional Approach; ALT-CI: alternative Control/Impact. Temporal comparator codes: Existing_Hydro_alt flow: existing HPP where one flow magnitude *Before* is compared to a new level *After* intervention; Existing_Hydro_base flow: existing HPP where one base flow magnitude *Before* is compared to a new base flow level *After* intervention; New_Hydro: flow prior to the installation of a new HPP. Spatial comparator: No_Hydro: a different nearby waterbody with no HPP; Upstream: upstream conditions in unmodified sections of the study waterbody; Up+Downstream: both up and downstream unmodified sections of the study waterbody; ALT_Hydro: a different nearby waterbody with HPP operating at a different but unmodified flow magnitude.

5.2.5 Outcomes

Fish were sampled with a variety of methods with many studies using more than one type of sampling method. This resulted in more cases than the total number of studies (143 cases from 133 studies). The most used method was gear (e.g., electrofishing, gill-, fyke-, seine-netting, trapping) (94 cases). Other methods included: visual techniques (nine cases), angling (five cases), mark-recapture (three cases) or a combination of techniques (19). No study used telemetry to sample fish and 12 cases used other techniques (i.e., historical or commerical catch data, hydroacoustics). Because some studies sampled with more than one type of method for different species/life stages, the total number of cases of sampling method exceeds the total number of studies when considering method by fish response (167 cases from 133 studies). Both fish abundance and biomass were sampled primarily with gear (113 cases) with the next most common method being some combination of techniques (32 cases).

Several studies conducted sampling during more than one season (53/133 studies), resulting in more cases than the total number of studies (174 cases). Of studies that sampled in more than one season 47% sampled in three or more. A total of 62 studies reported data for a single sampling season, with 29 studies sampling in summer, 25 sampling in fall, five sampling in spring and three studies sampled in winter only. Summer was sampled in the intervention (i.e., intervention site or *After* period) in 50% of all cases (considering cases where sampling was conducted in single and multiple seasons together), followed by fall (43%), spring (25%) and winter (11%). Four *BA* studies had mismatched sampling in the before and after period (i.e., sampled in summer in the *After* period, but fall in the *Before* period), while 13 *BA* studies had partial mismatch (i.e., one or more seasons matched, but other seasons were also present in the *Before* but not the *After* period).

Because studies often recorded more than one fish response (i.e., abundance and biomass in the same study) or more than one outcome metric (e.g., abundance and density), there are more outcome cases than studies (269 cases from 133 studies). Most cases considered abundance outcomes (82%) including the following metrics: (i) abundance, (ii) density, (iii) CPUE, (iv) number of eggs and (v) presence/absence. Abundance (42%) and density (40%) metrics accounted for the majority of abundance outcomes (Fig. 13). Biomass outcomes (including the metrics of biomass and yield) accounted for 18% of all cases, with only a single case considering yield (Fig. 13).

Many studies included outcomes for more than one life stage (i.e., adult, juvenile and larvae reported separately within the same study), resulting in more cases of life stages than the total number of studies. No life stage was reported for 119 of the 269 cases. Of those studies that reported life stage, the majority reported individual life stages separately (109 cases). The most commonly reported life stage for the abundance outcome was age-0 fish (28 cases), while the most commonly reported life stage for the biomass outcome was juveniles and age-0 fish (four cases each) (Fig. 13).

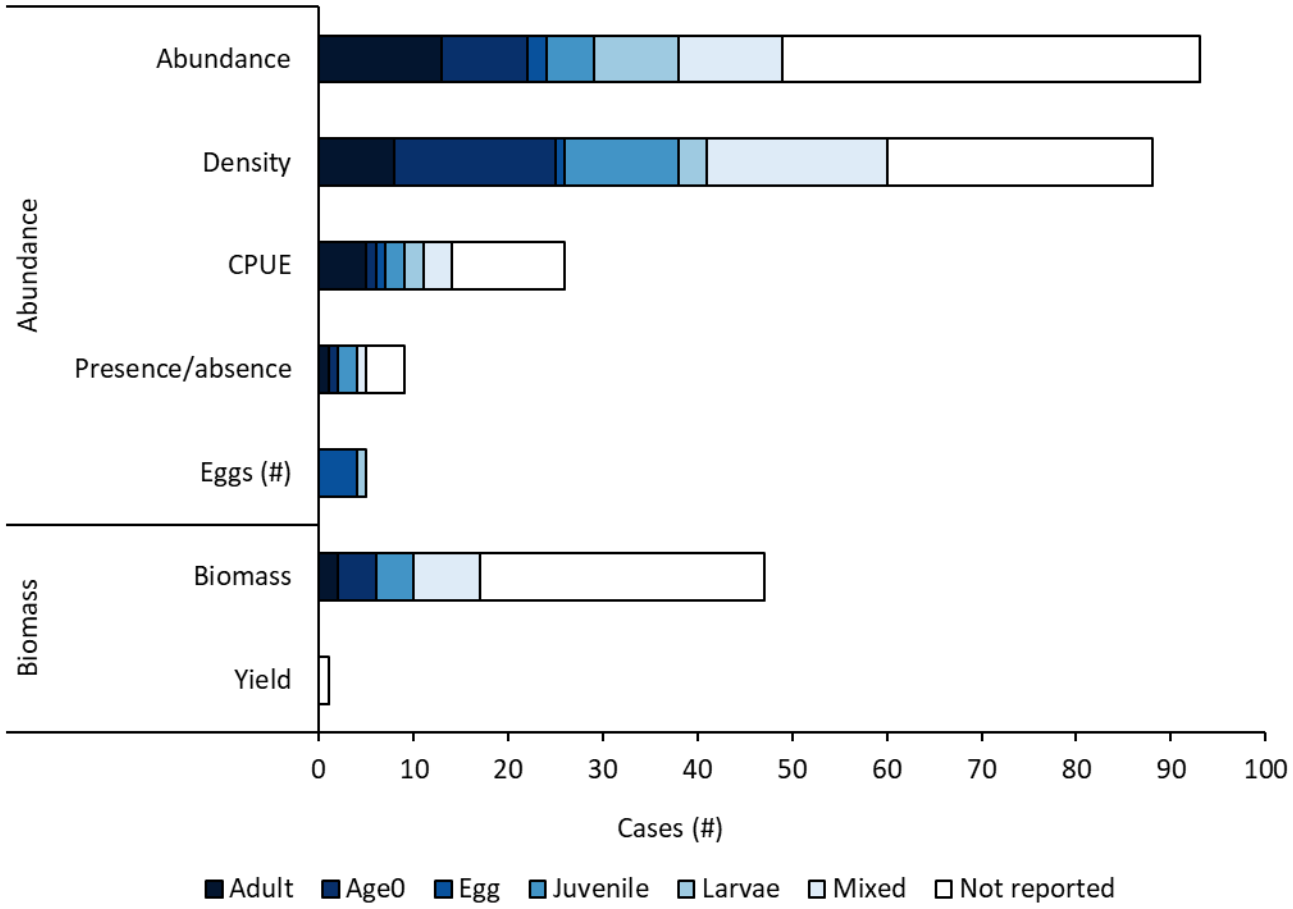

Fig. 13. Frequency of reported fish outcomes and life stage. Note: several studies reported more than one outcome and life stage separately. Mixed life stages include any combination of other life stages.

5.3 Quantitative synthesis

5.3.1 Description of the data

Of the 133 studies (from 103 articles) included in the narrative synthesis, 58 studies (from 46 articles) with 268 datasets, after aggregation, were included in our quantitative synthesis dataset (Appendix 10). Of these, 22 studies (91 datasets after aggregation) were used to analyze *Control/Impact* studies, while 37 studies (165 datasets after aggregation) were used in analyzing *Before/After* study designs. A single study had both a *CI* and *BA* component (i.e., an abstraction reach and outflow reach were included in a *BACI* design; the abstraction reach had fewer

temporal replicates and was converted to a *CI* design, while the outflow reach with fewer spatial replicates was converted to a *BA* design to maximize replication), so the total number studies (59) is greater than the actual number of studies included (58). We combined *CI*, *ALT-CI*, *RCA* and *BACI* converted to *CI* in analyses of *Control/Impact* studies and combined *BA*, deficient *BA*, incomplete *BACI* converted to *BA* and *BACI* converted to *BA* study designs in analysis of *Before/After* studies. We intended to analyze the potential for a time lag in *CI* designs when studies reported >1 post-treatment years, but there was insufficient sample size to allow for such an analysis; 3 studies (12 datasets) included more than one year of sampling separately, ranging from 2-4 years. These datasets were removed for all subsequent *CI* analyses and analyses proceeded with datasets for the first year post-treatment only.

Of the datasets included in the quantitative synthesis (256), 130 had 'Medium' overall study validity (51%), while 126 had 'Low' overall study validity (49%). Datasets used to analyze *CI* study designs were evenly distributed between 'Low' (44, 48%) and 'Medium' (47, 52%) validity, as were datasets used in the *BA* analyses (82 'Low' and 83 'Medium').

Datasets included for quantitative synthesis were primarily from North America (182 datasets), with most datasets being from the United States (105 of 182 datasets), followed by Canada (77 datasets). The next most represented region was Europe with 42 datasets from five countries. No datasets were included from South America or Eurasia. All included datasets occurred in river environments, although two datasets had sites in both riverine and estuarine environments.

A total of 32 hydropower facilities/dams were included in the quantitative synthesis database. Of these, seven dams (abundance: 60 datasets; biomass: four datasets) were considered by *CI* studies (Fig. 14), although an additional 26 datasets (abundance: 17 datasets; biomass: nine

datasets) were for studies considering multiple dams (i.e., as replicates) or did not report the names of the facilities considered. *BA* studies using within-year replication were conducted at seven dams/facilities (abundance: 37 datasets) (Fig. 14). *BA* studies with interannual replication considered 21 dams/facilities (abundance: 110 datasets; biomass: 17 datasets), although two datasets did not report sufficient information to determine dam/facility name (Fig. 14). Only three dams/facilities (Kinnaird Burn, Cow Green and Glen Canyon Dam) were considered in more than one study design. Of the nine dams/facilities with greater than ten datasets, only three dams were considered in more than one study (Glen Canyon Dam, Rupert Dam, Terzaghi Dam). Of the remaining dams/facilities, only two were considered by more than one study (two studies each).

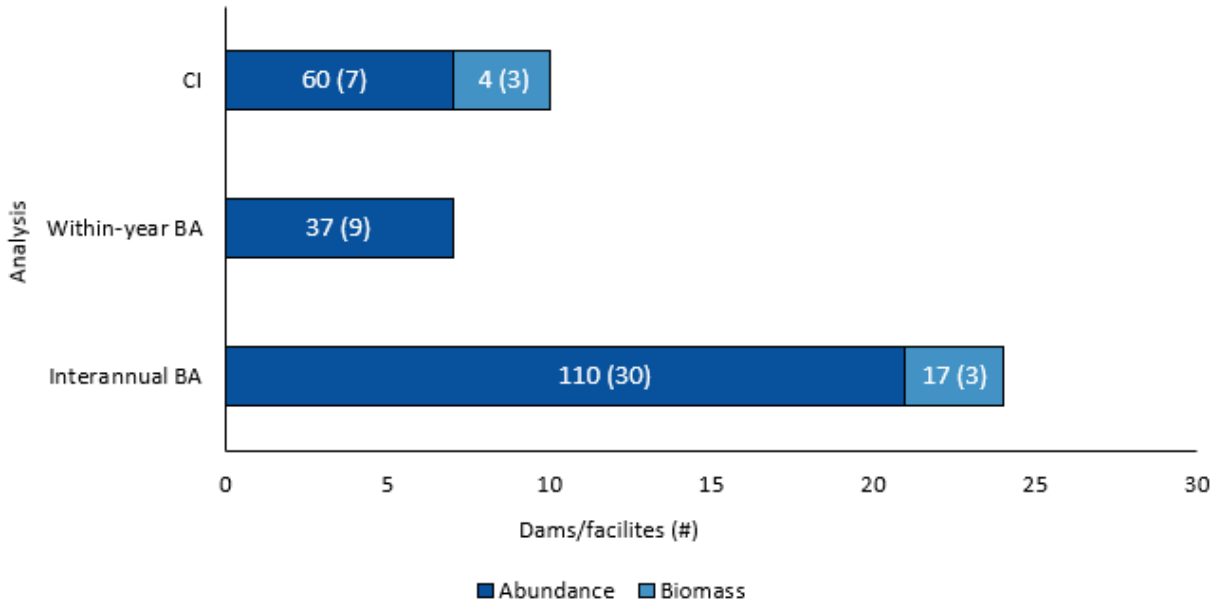

Fig. 14. Number of individual dams/facilities considered in each analysis type (i.e., *CI* or *BA*) and outcome (abundance and biomass). Numbers in columns indicate the number of datasets and studies (within brackets). Note that three dams/facilities were considered in more than one analysis (i.e., both in a Within-year and Interannual *BA*). An additional 28 datasets considered more than one dam/facility or did not provide sufficient information on location. *CI*: *Control/Impact*; *BA*: *Before/After*.

Of datasets that reported life stage (133/268), many reported more than one life stage, but combined quantitative results for life stages together (i.e., mixed life stages; 29%). Age-0 fish were the most frequently reported single age group (11%), followed by adults (5%), larvae (3%), number of eggs and juveniles (1% each). Half of datasets did not report the life stage of the studied fish.

For the quantitative synthesis database, five studies considered multiple species combined (i.e., quantitative data was pooled for all species; 12 datasets). All other studies either reported data for individual species (93 studies), or did not report taxonomic information (12 studies). Within datasets with individual species information (226), 98 species, from 57 genera and 27 families were evaluated for impacts of flow magnitude alterations. An additional 17 species were identified in multi-species groups but individual effect sizes could not be calculated for these species. The most commonly evaluated species were from the Salmonidae family, including *Oncorhynchus mykiss* (24 datasets), *Salmo trutta* (24 datasets), *Oncorhynchus kisutch* (10 datasets), *Salmo salar* (10 datasets) and *Oncorhynchus tshawytscha* (eight datasets). *Cottus gobio* was also evaluated in eight datasets and all other species were evaluated in fewer than eight datasets each.

Information on dam size (i.e., high, low or very low head height) and operation (peaking, run-of-river and storage) was available in 192/256 datasets. An additional 15 datasets had information on only operation, while 49 datasets had information only on dam size. The most common head height was high head (89%), followed by low head (3%) and very low head (2%). Storage operations were the most common operational regime (40%), followed by peaking (25%) and run-of-river (15%).

Of the 256 datasets, 226 datasets reported fish abundance outcomes, including 112 datasets of fish abundance, 78 of density, 34 of CPUE and two datasets that considered numbers of eggs. The remaining 30 datasets reported fish biomass (Table 6). Both *CI* and *BA* study design datasets primarily reported fish abundance (86% and 90% of datasets, respectively). Most datasets that described an alteration in flow magnitude evaluated the impact of increases in flow magnitude (40%). Of these, 38 datasets considered increases in base flow or increases to at least two flow magnitude elements (e.g., peak flow, base flow, average discharge or short-term variation; 48 datasets). The remaining datasets reported increases in average discharge (3 datasets) or short-term variation (4 datasets). Of the 61 datasets that reported decreases in flow magnitude, 37 reported alterations in average discharge, six reported changes in short-term variation, and one reported alterations to base flow. The remaining 17 datasets reported decreases in two or more flow magnitude elements. A minority of datasets reported both increases and decreases to flow magnitude elements (three datasets) and 90 datasets did not report sufficient information to assign a flow magnitude or flow element alteration.

Table 6. The number of datasets for the two different outcomes by flow magnitude element alterations (refer to Table 4 for definitions). *Multiple* (#) indicates that more than one element was altered and the number of elements that were either increased or decreased. *Unclear* indicates that although a change in flow magnitude was specified, the element considered was not clear, while *Unspecified* indicates that an alteration in flow magnitude was assumed due to the presence of a hydropower facility.

	Abundance	**Biomass**	**Total**
Peak Flow	9		9
Base Flow	29	10	39
Average Discharge	30	10	40
Short-term Variation	8	2	10
Multiple (2)	52	1	53
Multiple (3)	13	1	14
Multiple (4)	1		1
Unclear	9		9
Unspecified	75	6	81
Total	**226**	**30**	**256**

Gear based sampling was the most common sampling method used (85%), with all other sampling methods accounting for the remaining 15% (visual: 4%, other: 2%, angling: <1%, multiple 9% of datasets). Sampling occurred in all seasons, with 52% of datasets having sampling in more than one season, normally summer and fall (21%). When samples occurred in a single season, the most common was fall (25%) followed by summer (15%) and spring (5%). Fall was sampled in 193 of all datasets, summer in 173, spring in 93 and winter in only 30 datasets (but never without some other season also being sampled). Because more than 50% of datasets included more than one season, the total number of datasets evaluating seasons exceeds the total number of datasets (256). First sampling often occurred within the first year of the intervention (59% of datasets). Few datasets reported sampling beginning after the first year of the intervention (8% started between two and four years after); however, 12% of datasets reported that sampling began >4 years post-intervention. Additionally, 21% of studies did not report when post-intervention sampling was started. The range of monitoring duration for all datasets was between less than one and 36 years. Most datasets reported monitoring for three or fewer years (72%, with 21% of these monitoring for ≤1 year), while less than 5% of datasets reported greater than 10 years. A single study reported greater than 20 years of monitoring.

5.3.2 Global meta-analyses – Control/Impact studies

Abundance

The overall mean weighted effect size of *CI* studies for abundance was -0.001 (95% CI -0.35, 0.34; $k = 77$, $p = 0.997$; Table 7, Appendix 11: Fig. S1.), suggesting changes in flow magnitude did not significantly affect fish abundance. Over half of effect sizes were positive (i.e., $g>0$; 43 of 77), suggesting that changes in flow magnitude positively impacted fish abundance (i.e., abundance was higher in intervention sites than in control sites) with the remaining showing

neutral or negative responses (i.e., $g \leq 0$) to changes in flow magnitude; however, most of the individual effect sizes were not statistically significant, having confidence intervals that overlapped zero (70 out 77 effect sizes) (see forest plot Appendix 11: Fig. S1). The Q test for heterogeneity suggested that there was significant heterogeneity between effect sizes (Q = 104.05, p = 0.018). Funnel plots of asymmetry suggested possible evidence of publication bias towards larger studies showing positive effects of flow magnitude change (funnel plot Appendix 11: Fig. S2). Interestingly, this evidence of bias appears in the grey literature, not the commercially published literature. The failsafe number (N = 0) was not greater than 5k * 10 [(5 * 77 + 10) = 395], suggesting the results from the random effects model may not be robust against potential publication bias. The Cook's distance plot (see Appendix 11: Fig. S3) indicates a few influential effect sizes, with one outlier of concern (i.e., a single dataset with a comparatively large sample size relative to the other effect sizes). However, removing this single effect size from the random effects model yielded a similar result to the meta-analysis using all datasets, albeit a more positive overall effect size trend [Hedge's g = 0.06 (95% CI -0.30, 0.42; k = 76, p= 0.745)] (see forest plot Appendix 11: Fig S4) and had no impact on funnel plot asymmetry (see Appendix 11: Fig S5) or the failsafe number.

Table 7. Summary statistics from the main analyses of abundance and biomass including for *Control/Impact*, within-year and interannual *Before/After* studies, and from taxonomic-specific analyses. Hedges' *g* is the standardized mean difference effect size. CI: 95% confidence interval. Bold indicates significant effect ($p<0.05$); * indicates marginally significant effect ($p<0.1$). *k*: number of effect sizes.

Analysis	**Standardized mean difference (Hedges' *g*)**
Control/Impact	
(A) Global meta-analyses	
Abundance ($k = 77$)	-0.001 (95% CI -0.35, 0.34; $p = 0.997$)
Biomass ($k = 13$)	-0.16 (95% CI -0.62, 0.30; $p = 0.490$)
(B) Taxonomic analyses	
Abundance	
Catostomidae ($k = 4$)	-0.46 (95% CI -1.79, 0.88; $p = 0.503$)
Centrarchidae ($k = 6$)	0.52 (95% CI -0.83, 1.14; $p = 0.753$)
Cyprinidae ($k = 19$)	-0.09 (95% CI -1.20, 1.03; $p = 0.878$)
Percidae ($k = 4$)	0.32 (95% CI -0.84, 1.49; p=0.590)
Salmonidae ($k = 5$)	0.18 (95% CI -0.58, 0.93; $p = 0.646$)
Biomass	N/A
Before/After -within year	
(C) Global meta-analyses	
Abundance – year 1($k = 19$)	0.25 (95% CI -0.04, 0.56; $p = 0.091$)*
Abundance – year 2 ($k = 5$)	0.67 (95% CI -0.09, 1.43; $p = 0.084$)*
Abundance – year 3 ($k = 4$)	0.31 (95% CI -0.62, 1.23; $p = 0.516$)
Abundance – year 4 ($k = 3$)	0.20 (95% CI -0.80, 1.20; $p = 0.697$)
Abundance – year 1-4 ($k = 19$)	0.25 (95% CI -0.02, 0.52; $p = 0.072$)*
Biomass	N/A
(D) Taxonomic analysis	
Abundance	
Salmonidae ($k = 4$)	0.81 (95% CI -0.15, 1.76; k = 4, p = 0.099*)
Biomass	N/A
Before/after – interannual	
(E) Global meta-analysis	
Abundance ($k = 112$)	0.19 (95% CI -0.23, 0.61; $p = 0.374$)
Biomass ($k = 17$)	0.46 (95% CI -0.24, 1.15; $p = 0.196$)
(F) Taxonomic analysis	
Abundance	
Acipenseridae ($k = 5$)	0.42 (95% CI -1.98, 2.81; $p = 0.733$)
Anguillidae ($k = 5$)	-0.45 (95% CI -1.44, 0.55; $p = 0.379$)
Catostomidae ($k = 8$)	-0.38 (95% CI -1.84, 1.07; $p = 0.606$)
Centrarchidae (k=5)	7.14 (95% CI -5.68, 19.96; $p = 0.275$)
Cottidae ($k = 5$)	**1.34 (95% CI 0.39, 2.29; $p = 0.006$)**
Cyprinidae (k=15)	-1.18 (95% CI -3.26, 0.90; $p = 0.266$)
Esocidae ($k = 3$)	0.37 (95% CI -0.36, 1.10; $p = 0.325$)
Ictaluridae ($k = 3$)	6.69 (95% CI -6.86, 20.24; $p = 0.333$)
Salmonidae ($k = 59$)	**0.45 (95% CI 0.25, 0.65; $p < 0.0001$)**
Biomass	
Salmonidae ($k = 11$)	0.52 (95% CI -0.38, 1.43; $p = 0.258$)

N/A: Unable to conduct moderator analyses due to insufficient sample size or variability. A decrease in fish abundance from alterations to flow magnitude due to HPP compared to control groups is indicated by a value <0 for Hedges' *g*.

The sensitivity analyses for both medium validity studies and studies with true replication showed a more negative effect of flow magnitude changes on fish abundance compared to the overall meta-analysis. The difference in the relative magnitude of effect sizes for these analyses suggests that the results may not be fully robust to the inclusion of studies with low validity or pseudoreplication, but the effect sizes were non-significant (Table 8). Sensitivity analyses of studies where no stocking influenced outcomes indicated a slight increase in the effect size relative to the overall meta-analysis, as did the analyses of studies that specified flow magnitude specifically (i.e., removing studies reporting comparisons of regulated to unregulated systems without a measure of flow magnitude); however, the results are comparable to the overall meta-analysis and the effect sizes are non-significant. This indicates the results may be robust to the inclusion of studies that: (i) may have been impacted by stocking, or (ii) studies that did not specify a flow magnitude (Table 8). The sensitivity analysis based only on studies comparing a single intervention to a single comparator showed a positive effect of flow magnitude change on fish abundance compared to the overall meta-analysis, but the relative magnitude of the effect sizes was comparable. Similarly, the sensitivity analysis for studies where imputation of variances was necessary had a comparable effect size to the overall meta-analysis (Table 8). In both cases, the summary effect was non-significant, again indicating that the overall meta-analysis result may be robust to the inclusion of studies comparing multiple interventions with a single comparator, or where imputation was used.

Table 8. Summary statistics of applicable sensitivity analyses for *CI* study designs and abundance. Statistical significance at $p<0.05$.

Analysis	**Standardized mean difference (Hedge's *g*)**
Global analysis (k = 77)	-0.001 (95% CI -0.35, 0.34; p = 0.997)
Medium Validity (k = 39)	-0.18 (95% CI -0.60, 0.23; p= 0.385)
Without imputation (k = 75)	0.02 (95% CI -0.34, 0.37; p = 0.923)
Without stocking (k = 66)	-0.03 (95% CI -0.48, 0.43; p=0.900)
True replication (k = 11)	-0.10 (95% CI -0.55, 0.35; p = 0.6721)
Single comparator/single intervention (k = 69)	0.005 (95% CI -0.47, 0.48; p = 0.985)
Reported flow magnitude components (k = 44)	-0.04 (95% CI -0.47, 0.39; p = 0.858)

Note, a decrease in the abundance of fish from alterations to flow magnitude due to HPP compared to control groups is indicated by a value <0 for Hedges' *g*. CI: 95% confidence interval. *k*: number of effect sizes.

Biomass

The overall weighted mean effect size for biomass was -0.16 (95% CI -0.62, 0.30; k = 13, p = 0.490), suggesting that changes in flow magnitude negatively impacted fish biomass (i.e., biomass was lower in intervention sites than in control sites), but the response was not significant (Table 7). Most effect sizes were negative (i.e., $g<0$; 7 out of 13); however, only one effect size was statistically significant (see forest plot Appendix 11: Fig. S6). The Q test of heterogeneity suggested that there was moderate statistical significance in heterogeneity between effect sizes (Q = 20.94, p = 0.051). The funnel plot of asymmetry did not suggest an obvious pattern of publication bias; however, it is difficult to determine asymmetry with this small number of studies (k=13; Appendix 11: Fig. S7). The failsafe number (N=0) was not greater than $5k+10$ [(5*13+10)=75], suggesting the results from the random effects model may not be robust against potential publication bias.

Sensitivity analysis based on only medium validity studies showed a more positive effect of changes to flow magnitude on fish biomass compared to the overall meta-analysis (Table 9). The summary effect was not significant, suggesting that the result of the overall meta-analysis

may be robust to differences in study validity. All other sensitivity analyses for biomass and *CI* study designs had slightly larger, more negative effect sizes; however, results were comparable to the overall meta-analysis (Table 9) and no effect sizes were significant. This indicates that results may be robust against the inclusion of studies with imputation, potential impacts of stocking, pseudoreplication, multiple interventions compared to a single comparator, and studies that do not report flow magnitude components (i.e., studies reporting regulated vs unregulated systems without also reporting a flow magnitude component, or reporting multiple flow magnitude components).

Table 9. Summary statistics of applicable sensitivity analyses for *CI* study designs and biomass. Statistical significance at $p<0.05$.

Analysis	**Standardized mean difference (Hedge's *g*)**
Global analysis (k = 13)	-0.16 (95% CI -0.62, 0.30; p = 0.489)
Medium Validity (k = 7)	0.07 (95% CI -0.46, 0.61; p = 0.789)
Without imputation (k = 12)	-0.16 (95% CI -0.64, 0.32; p = 0.508)
Without stocking (k = 10)	-0.25 (95% CI -0.79, 0.29; p = 0.366)
True replication (k = 8)	-0.26 (95% CI -0.83, 0.31; p = 0.369)
Single comparator/single intervention (k = 5)	-0.33 (95% CI -1.20, 0.54; p = 0.458)
Reported flow magnitude components (k = 10)	-0.25 (95% CI -0.79, 0.29; p = 0.366)

Note, a decrease in the abundance of fish from alterations to flow magnitude due to HPP compared to control groups is indicated by a value <0 for Hedges' *g*. CI: 95% confidence interval. *k*: number of effect sizes.

5.3.3 Effects of moderators – Control/Impact studies

Abundance

When addressing potential reasons for heterogeneity in the results, there was only sufficient sample size (i.e., ≥3 datasets from ≥2 studies) to address effect-modifying factors for *CI* study designs and abundance. Additionally, there were too few effect sizes (in sufficiently different categorical levels) to allow meaningful analysis of waterbody type (i.e., all included studies

occurred in rivers) or 'life stage' (only Mixed and Not reported life stages were present). We present the results of all other moderator analyses here, and summarize the outputs of conducted moderator analyses for *CI* study designs in Table 10.

Due to sample size, we either combined or dropped levels within moderators as follows:

- Dam size (i.e., head height): (i) High; (ii) Low + Very-low; (iii) Unclear.
- Hydropower operational regime: (i) Peaking; (ii) Run-of-river; (iii) Storage.
- Direction of flow magnitude alteration: there was only sufficient sample size to consider two interventions: (i) Increase; (ii) Unspecified. We collapsed any increase of average, peak or base flow and short-term variation into 'Increase' and any decrease in average, peak or base flow and short-term variation into 'Decrease'. However, the single effect size with a decrease in flow magnitude was not included during analysis.
- Alterations to other flow components: (i) Yes (i.e., alterations present); (ii) No; (iii) Unclear+Not reported.
- Sampling methods: (i) Gear; (ii) Multiple (i.e., any combination of methods). Four datasets from a single study using visual sampling methods were not included during analysis.
- Sampling seasons: (i) Summer; (ii) Fall; (iii) Summer+Fall; (iv) 'Spring+Summer+Fall'+All seasons.
- Type of comparator (spatial): (i) Upstream (i.e., upstream of dam/facility); (ii) No hydro (i.e., separate but similar waterbodies without HPP). Because it was not possible to combine comparator types, the single effect size with alternative hydropower as a comparator was not included in the analysis. No effect sizes with upstream and downstream comparators were included in the analysis.

- Time since intervention (years): (i) ≤1 year; (ii) >4 years; (iii) Unclear+Not reported.
- Monitoring duration (years): (i) ≤1 year; (ii) 2 years; (iii) 3 years; (iv) ≥ 5 years. Due to sample size, we considered monitoring duration a categorical variable in *CI* studies.

For all moderators considered, we found no detectable effect on the average effect size (Table 10). Additionally, most moderators were highly correlated (see results of Pearson chi-square test; Appendix 11: Table S1).

Table 10. Summary results of meta-analyses using subsets of fish abundance effect sizes for *Control/Impact* studies, testing the influence of the given moderator variable. Bold indicates statistical significance ($p<0.05$).

Moderator	***k***	***Q* statistic (*p*-value)**	**Q_M (*p*-value)**	**Q_E (*p*-value)**
Waterbody type	77	N/A		
Dam size				
Unmoderated model	77	**104.05 (*p*=0.018)**	-	-
Dam size	77	-	0.52 (*p*=0.773)	**98.22 (*p*=0.031)**
Hydropower operational regime	77	**104.05 (*p*=0.018)**	-	-
Unmoderated model	77	-	2.85 (*p*=0.241)	90.12 (*p*=0.098)*
Operational regime				
Direction of flow magnitude alteration				
Unmoderated model	76	**103.24 (*p* = 0.017)**	-	-
Flow alteration	76	-	0.36 (*p*=0.551)	92.12 (*p*=0.076)*
Alterations to other flow components				
Unmoderated model	77	**104.05 (*p*=0.018)**	-	-
Other components	77	-	0.27 (*p*=0.873)	**99.31 (*p*=0.027)**
Sampling methods				
Unmoderated model	73	**96.90 (*p*= 0.028)**	-	-
Sampling technique	73	-	0.07 (*p*=0.789)	**96.29 (*p*=0.025)**
Sampling season				
Unmoderated model	77	**104.05 (*p*=0.018)**	-	-
Sampling season	77	-	2.21 (*p*=0.531)	**94.90 (*p*=0.044)**
Type of comparator (spatial)	76	**97.46 (*p*=0.042)**	-	-
Unmoderated model	76	-	2.04 (*p*=0.153)	86.01 (*p*=0.161)
Comparator site				
Time since intervention				
Unmoderated model	77	**104.05 (*p* = 0.018)**	-	-
Time since intervention	77	-	1.64 (*p*=0.441)	88.06 (*p*=0.126)
Monitoring duration				
Unmoderated model	75	**97.46 (*p*=0.035)**	-	-
Monitoring duration	75	-	3.88 (*p*=0.143)	83.97 (*p*=0.158)
Life stage	77	N/A		

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. * Significance at $p<0.1$. N/A: unable to assess moderator due to insufficient sample size or lack of variation.

Biomass

There were too few effect sizes within biomass and *CI* studies (k=13) to allow for meaningful analysis of potential effect modifiers.

5.3.4 Global meta-analyses – Before/After studies

Within-year Before/After studies

Abundance

The overall mean weighted effect size of within-year *BA* studies, considering only effect sizes for abundance in post-intervention year-1 was 0.25 (95% CI -0.04, 0.56; k = 19, p = 0.091), suggesting change in flow magnitude had a moderately significant overall positive effect on fish abundance (Table 7C, Fig. 14; Appendix 11: Fig. S8.). However, the sample size was small and there was a single study (with a single dataset) with a much larger, significant positive effect size, which may have disproportionately impacted the mean effect size for increasing abundance [non-native, established *Oncorhynchus mykiss* (Avery et al. 2015); see Appendix 11: forest plot Fig. S8 and Cook's distance Fig. S10]. Most other effect sizes were also positive (i.e., g>0; 13 of 19), with the remaining showing neutral or negative responses (i.e., $g \leq 0$) to changes in flow magnitude. All other effect sizes were not statistically significant and had confidence intervals that overlapped zero (18 out 19 effect sizes) (see forest plot Appendix 11: Fig. S8). The Q test for heterogeneity did not suggest significant heterogeneity between effect sizes (Q = 15.65, p = 0.617). There was no obvious indication of publication bias from the funnel plot, although it was difficult to determine asymmetry with this small number of studies (Appendix 11: Fig S9). However, the failsafe number was zero suggesting the results from the random effects model may not be robust against potential publication bias.

Sensitivity analyses for within-year *BA* studies were conducted only for year-1 post-intervention datasets. Sensitivity analyses to determine the effect of the inclusion of studies with imputation of missing variances, single comparators and multiple interventions, deficient BA or BACI studies, and inclusion of outflow regions were not conducted because datasets with these features were not present for within-year *BA* studies. Sensitivity analysis of medium validity studies showed a slightly smaller, but still positive effect size of flow magnitude changes on fish abundance, but this overall response was no longer significant (possibly due to a smaller sample size for medium validity studies only), compared to the overall meta-analysis [Hedges' $g = 0.14$ (95% CI -0.17, 0.44; $k = 14$, $p = 0.386$). The small difference in average effect sizes between the two models suggests that the results may be robust to inclusion of studies with low validity. The sensitivity analysis of studies where stocking clearly did not occur showed a statistically significant and slightly more positive mean effect size than the overall meta-analysis [Hedge's $g = 0.35$ (95% CI 0.02, 0.69; $k = 17$, $p = 0.040$)], indicating that the results may not be robust against the inclusion of studies that were potentially impacted by stocking. Similarly, the sensitivity analysis of only studies that reported flow magnitude components showed a statistically significant and slightly more positive mean effect size than the overall meta-analysis [Hedge's $g = 0.35$ (95% CI 0.02, 0.69; $k = 17$, $p = 0.040$)]. However, in both cases the same two effect sizes were removed. Given the small sample size and the removal of two of the six negative effect sizes, it is possible that the response of the model to the removal of these two effect sizes may not be associated with potential stocking or the lack of specific flow magnitude description *per se*, but with some other aspect of the study. It is also important to note that the mean effect sizes for the model with and without potential impacts of stocking, and therefore flow magnitude elements, were within the confidence intervals of the two models (see Appendix

11: forest plots with stocking/no reported flow magnitude Fig. S8, without stocking/no reported flow magnitude Fig. S15).

To investigate the potential impact of a time-lag in within-year fish responses to changes in magnitude, we compared the effect sizes for subsequent years of sampling post-intervention to that of the overall mean weighted effect size for year-1 datasets. The overall mean weighted effect size of post-intervention year-2 for abundance was moderately significant 0.67 (95% CI -0.09, 1.43; $k = 5$, $p = 0.084$) indicating that a slight increase in abundance may occur after two years of monitoring (Fig. S11). However, the sample size was quite small and one study had a significant effect size (non-native, established *Oncorhynchus mykiss*; Korman et al. 2011) which may have had a disproportionate effect on the overall mean effect size. The overall mean weighted effect sizes after post-intervention year-2 decreased. By post-intervention year-4, the effect size had returned to a value similar to, but slightly lower, than those of post-intervention year-1 [post-intervention year-3: Hedge's $g = 0.31$ (95% CI -0.62, 1.23; $k = 4$, $p = 0.516$); post-intervention year-4: Hedge's $g = 0.20$ (95% CI -0.80, 1.20; $k = 3$, $p = 0.697$)] (Fig. 14; Appendix 11: Fig. S12-13). Of the 15 species present in year-1, only three species were present in all four post-intervention years (*Cottus gobio*, *Oncorhynchus mykiss*, *Salmo trutta*). Year-2 had one species not present in year-3 and -4 (*Sinibotia superciliaris)*. When all post-intervention years (1-4) were aggregated, the resulting overall mean weighted effect size was very similar to that of year-1 alone [Hedge's $g = 0.25$ (95% CI -0.02, 0.52; $k = 19$, $p = 0.072$)] (Fig. 15), with a moderately significant effect of changes in flow magnitude on fish abundance (see forest plot Appendix 11: Fig. S14). For each post intervention year separately, and for aggregated years 1-4 post-intervention, the Q test for heterogeneity suggested that there was no significant

heterogeneity between effect sizes. Due to small sample sizes, it was not possible to investigate the effect of moderators for within-year *BA* studies.

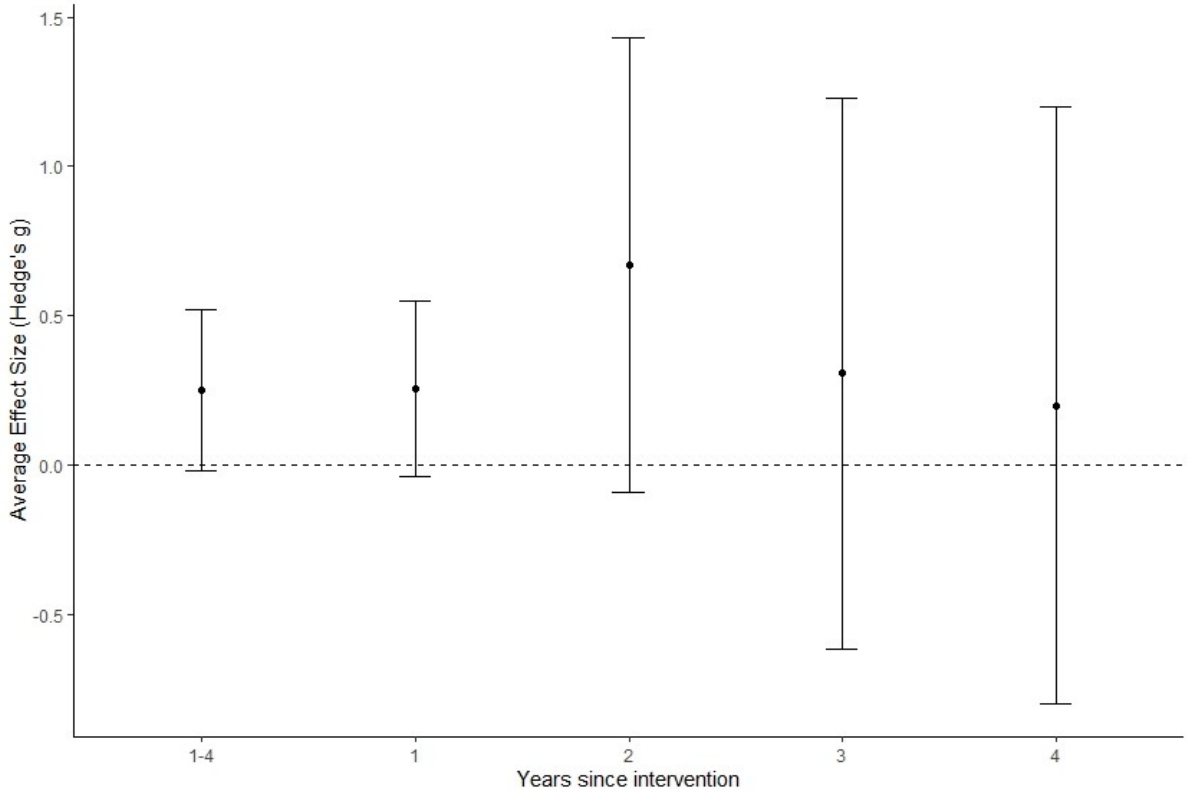

Fig. 15. Comparison of overall average effect size for within-year *BA* studies one (*k*=19), two (*k* = 5), three (*k* = 4) and four (*k* = 3) years post- intervention and when after years 1-4 were aggregated (year 1-4). Models were developed for each of the first four years after a change in flow magnitude [i.e., comparing the most recent or only *Before* year with *After* year-1 only, *After* year-2 only, *After* year-3 only, and *After* year-4 only, as well as the average of years 1-4 after a change in flow magnitude. *k*: number of effect sizes.

Biomass

No study using within-year temporal replication considered biomass as an outcome metric.

Therefore, quantitative analysis for this subset of *Before/After* studies only considers abundance.

Interannual Before/After studies

Abundance

The overall mean weighted effect size for abundance when considering interannual *BA* studies was 0.19 (95% CI -0.23, 0.61; $k = 112$, $p = 0.374$), indicating that alterations in flow magnitude had a slight positive, but non-significant, overall effect on fish abundance (see forest plot Appendix 13: Fig., S16). Most effect sizes were positive (i.e., $g>0$; 77 of 112), with the remaining 35 effect sizes showing neutral or negative responses (i.e., $g\leq 0$) to alterations in flow magnitude. Most individual effect sizes were not statistically significant with confidence intervals overlapping zero (89 of 112; Appendix 13: Fig. S16), although 10 datasets were significant negative, and 13 datasets were significant positive. The Q test for heterogeneity suggested that there was significant heterogeneity between effect sizes ($Q = 421.12$, $p<0.0001$) that could be explored using mixed effects models (see section "Effects of Modifiers – Interannual *BA* studies" below). The funnel plot of asymmetry suggests possible evidence of publication bias, especially in grey literature [i.e., as study sample size increased, the variance in effect sizes increased (see Appendix 11: Fig. S17)]. Also, the fail-safe number (N = 407) was less than 5k+10 [(5*112+10) = 570], suggesting the results of the random effects model may not be robust to publication bias. For this analysis, we assumed that any study with more than one intervention used the same comparator for all interventions [i.e., if two flow magnitude alterations (trials) are conducted, each can be compared to Trial 0 (the *Before* period), Trial 0 vs. Trial 1 and Trial 0 vs. Trial 2]; however, it is possible that each subsequent intervention could be compared to the previous intervention (i.e., Trial 0 vs. Trial 1, Trial 1 vs. Trial 2). When comparing these two options (i.e., changing the before period for two studies with multiple intervention periods; Appendix 11: Fig. S19), using different comparator periods did not greatly

alter the overall mean effect size [Hedge's g = 0.17 (95% CI -0.25, 0.59; k = 112, p = 0.4238)] and the direction of effect did not change (see Appendix 11: Fig. S20). We conducted all further analysis using the same comparator period.

Sensitivity analysis of studies without imputation of missing variance was not possible for interannual *BA* studies because no datasets required these calculations. All sensitivity analyses applicable to interannual *BA* study designs had similar results to the overall meta-analysis (Table 11) indicating that the results of the overall meta-analysis were robust to the inclusion of: (i) studies with low validity; (ii) studies where stocking may have impacted fish outcomes; (iii) studies with multiple interventions and a single comparator; (iv) studies that did not report flow magnitude components (i.e., *BACI* studies that compare before and after the start of regulation, in a regulated and unregulated stream, but do not state what alterations to flow magnitude occurred); (v) studies with deficient *BA* or *BACI* study designs; (vi) studies that averaged yearly averages in the *Before* and *After* periods (see Appendix 11: Fig 21-23 for comparisons of the effect sizes for different types of temporal data); and (vii) outfall zones.

Table 11. Summary statistics of applicable sensitivity analyses for interannual *BA* study designs and abundance. Statistical significance at $p<0.05$.

Analysis	**Standardized mean difference (Hedge's *g*)**
Global analysis (k = 112)	0.19 (95% CI -0.23, 0.61; p = 0.374),
Medium Validity (k = 40)	0.11 (95% CI -0.32, 0.54; p = 0.602)
Without imputation (k = 112)	N/A
Without stocking (k = 70)	0.19 (95% CI -0.41, 0.79; p = 0.535)
Single comparator/single intervention (k = 101)	0.17 (95% CI -0.29, 0.64; p = 0.464)
Reported flow magnitude components (k = 82)	0.09 (95% CI -0.37, 0.56; p = 0.694)
Without deficient BA or BACI (k = 108)	0.17 (95% CI -0.27, 0.61; p = 0.453)
Without averages of averages (k = 91)	0.15 (95% CI -0.41, 0.71; p = 0.595)
Inclusion of outfall zone (k = 103)	0.18 (95% CI -0.27, 0.62; p = 0.439

Note, a decrease in the abundance of fish from alterations to flow magnitude due to HPP compared to control groups is indicated by a value <0 for Hedges' *g*. CI: 95% confidence interval. *k*: number of effect sizes. N/A: unable to assess moderator due to insufficient sample size or lack of variation.

Biomass

The overall mean weighted effect size for biomass when considering interannual *BA* studies was 0.46 (95% CI -0.24, 1.15; k = 17, p = 0.196) suggesting an overall increase in fish abundance with alterations in flow magnitude compared to *Before* periods; however, the estimated overall response was not significant (see forest plot Appendix 13: Fig. S24). Most effect sizes were positive (10/17) while the remaining seven effect sizes were negative. The Q test for heterogeneity suggested that there was significant heterogeneity between effect sizes (Q = 28.37, p = 0.03) that could be explored using mixed effects meta-analysis models; however, given the small number of effect sizes, the influence of categorical moderators could not be assessed due to potential overparameterization, and was not possible for continuous moderators due to lack of variability in the datasets (i.e., there were too few years of data for each year and gaps between years were too large to allow effect meta-regression). The funnel plot for the random effects model did not show any obvious pattern of publication bias, but with the small sample size,

determination of asymmetry was difficult (note that no grey literature considered biomass for interannual *BA* studies; Appendix 13: Fig. S25). The failsafe number was two (less than the suggested 5k+10 [5*17+10 = 95], indicating the results of the random effect model for biomass may not be robust against publication bias. A single effect size was significant and positive, but no points were highly influential (see Cook's distance plot Appendix 11: Fig. S26). As with abundance, there was one study with multiple interventions that could have been compared to either the same comparator, or different comparator periods. Changing the comparator had little impact on the mean effect size (see Appendix 11: Fig. S27).

Sensitivity analysis was not conducted to assess the impact of including studies with imputation, studies that did not report flow magnitude components, deficient *BA* or *BACI* designs, or the inclusion of outflow regions because no datasets had these features. Sensitivity analysis of medium validity studies showed a slightly smaller, but still positive, effect of flow magnitude alterations on fish biomass compared to the overall meta-analysis, as did the sensitivity analysis of studies where no influence of stocking was present (Table 12). The effect sizes for these sensitivity analyses were not significant, indicating that the results of the random effects model may be robust to the inclusion of low validity studies or studies that may have been influenced by stocking. Sensitivity analysis of studies with only a single intervention (i.e., one alteration to flow during the study, rather than several alterations being compared to a single *Before* period) showed a larger, more positive effect size compared to the overall meta-analysis. Sensitivity analysis of studies that reported single data points or summed abundance per year, showed a larger, more positive effect size compared to the overall meta-analysis which included studies where yearly averages, averaged for the *Before* or *After* period (i.e., averages of averages) were used (Table 12). In both cases, the effect size for these sensitivity analyses were significant,

indicating that the results of the random effects model may not be robust to the inclusion of studies with more than one intervention, or that report averages of averages.

Table 12. Summary statistics of applicable sensitivity analyses for interannual *BA* study designs and biomass. Statistical significance at $p<0.05$.

Analysis	**Standardized mean difference (Hedge's *g*)**
Global analysis (k = 17)	0.46 (95% CI -0.24, 1.15; p = 0.196)
Medium Validity (k = 16)	0.34 (95% CI -0.46, 1.14; p = 0.403)
Without imputation (k = 17)	N/A
Without stocking (k = 16)	0.34 (95% CI -0.46, 1.14; p = 0.403)
Single comparator/single intervention (k = 8)	0.83 (95% CI 0.26, 1.41; p = 0.004)
Reported flow magnitude components (k = 17)	N/A
Without deficient BA or BACI (k = 17)	N/A
Without averages of averages (k = 7)	0.79 (95% CI 0.17, 1.41; p = 0.013)
Inclusion of outflow regions (k = 17)	N/A

Note, a decrease in the abundance of fish from alterations to flow magnitude due to HPP compared to control groups is indicated by a value <0 for Hedges' *g*. CI: 95% confidence interval. *k*: number of effect sizes. N/A: unable to assess moderator due to insufficient sample size or lack of variation.

5.3.5 Effects of moderators – Interannual Before/After studies

Abundance

To test for the influence of moderator variables on average fish responses to changes in flow magnitude, there were only sufficient sample sizes for interannual *BA* studies and fish abundance; there were too few effect sizes for abundance and biomass in the within-year *BA* studies and for biomass in the interannual *BA* studies to permit meaningful analyses. For all analyses, we present the main results of univariate mixed models, and summarize all model results in Table 13 and significant model results in Fig. 21. Because of significant correlation among most moderators and small sample sizes, we were unable to combine multiple moderators into single multi-variate models (see results of Pearson's χ^2 test; Appendix 11: Table S2).

However, the inclusion of these moderators left significant heterogeneity in all moderated models (Table 13), suggesting that interactions between moderators may be occurring, or other factors not captured by our analyses are influencing fish responses.

Waterbody type
There was no variation in waterbody type to investigate this moderator (i.e., all 112 datasets occurred in river systems).

Dam size
There was insufficient variation in dam size to investigate this moderator (i.e., all 112 datasets were from studies at high head dams).

Hydropower operational regime
We found no detectable effect of operational regime on average effect sizes (Table 13).

Direction of flow magnitude alteration
There were only sufficient sample sizes and variation to permit meaningful tests of the influence of grouped increases and decreases in flow magnitude, rather than increases and decreases in average discharge, peak flow, base flow and short-term variation separately (see Table 4 for definitions of flow magnitude intervention terms). The following levels within flow magnitude alterations were considered: (i) increase (combination of any increases in flow magnitude); (ii) decrease (combination of any decreases in flow magnitude); (iii) unclear (there was insufficient information to determine the overall increase/decrease in flow, but it was clear that flow magnitude had been altered); and (iv) unspecified (an alteration to flow magnitude was assumed but not explicitly stated). There was a statistically significant effect of the direction of flow magnitude on average fish abundance detected (Table 13), with studies including unclear alterations in flow associated with larger, positive effect sizes than those specifying decreases or increases in flow magnitude alterations (Fig. 16 and 21); although average effect size for unclear

alterations was only moderately significant and confidence intervals overlapped among all groups.

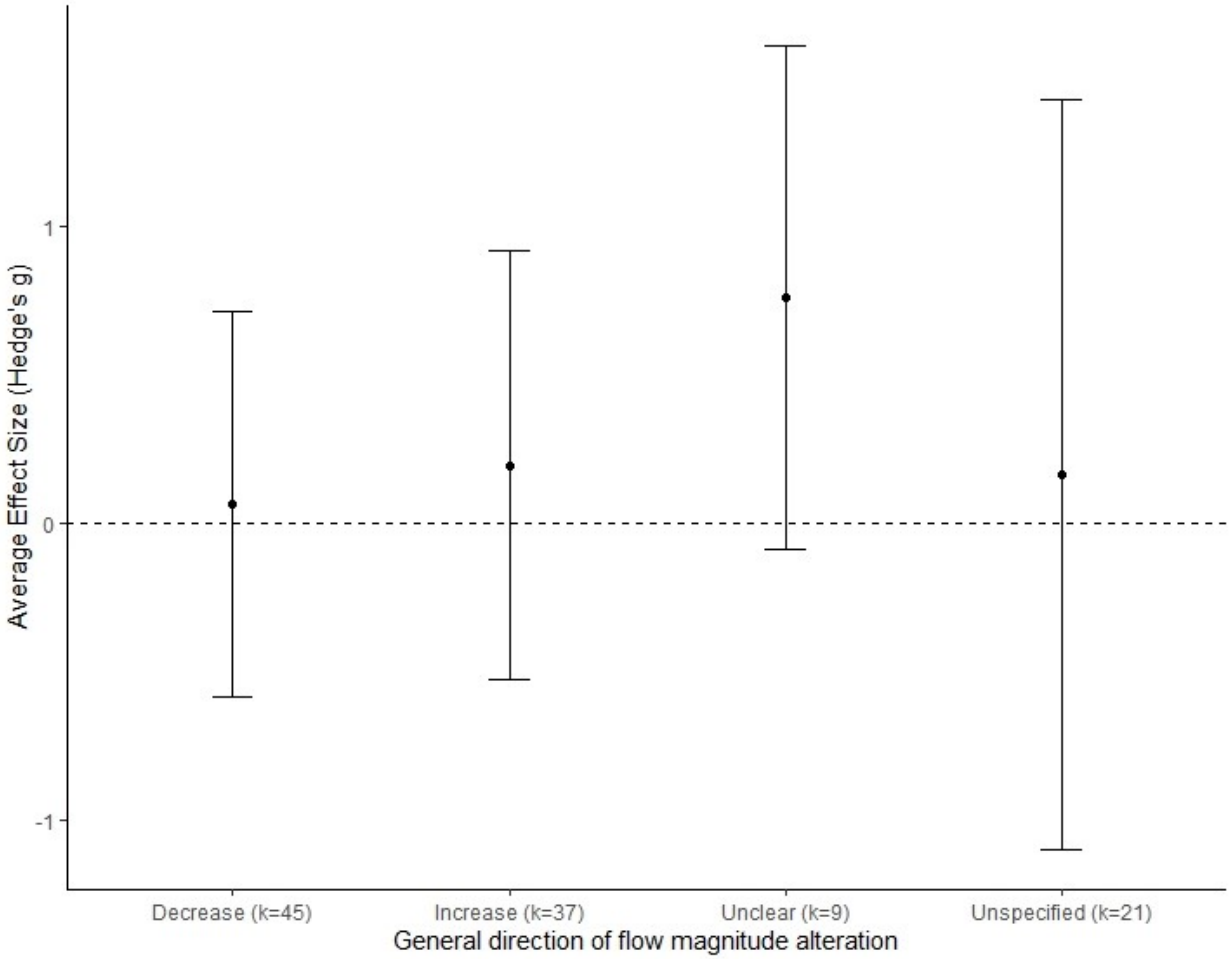

Fig. 16. Average effect size by intervention for fish abundance responses when considering interannual *BA* studies. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (above the dashed zero line) indicates that the abundance was higher in the *After* period (intervention) than in the *Before* period (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level). *Decrease*: flow magnitude in *After* period decreases in relation to *Before* period; *Increase*: flow magnitude in *After* period increases in relation to the *Before* period; *Unclear*: unclear description of flow magnitude alteration; *Unspecified*: no specified alteration to flow magnitude but change is assumed due to presence of hydropower facility (i.e., before and after the closure of a dam).

Alterations to other flow components

The presence of alterations to other flow components (i.e., frequency, duration, timing, rate of change, or surrogates of flow) was associated with average effect sizes (Table 13 and Fig. 21),

although the response of fish abundance varied among interventions. Studies lacking alterations to other flow components had the largest, most positive average effect size (i.e., higher abundance in the *After* period than the *Before* period), while studies that did not report whether other flow components were altered had the most negative average effect size (Fig. 17 and 21). Studies that did report the presence of alterations to other flow components had a slightly negative, but non-significant average effect size (Fig. 17).

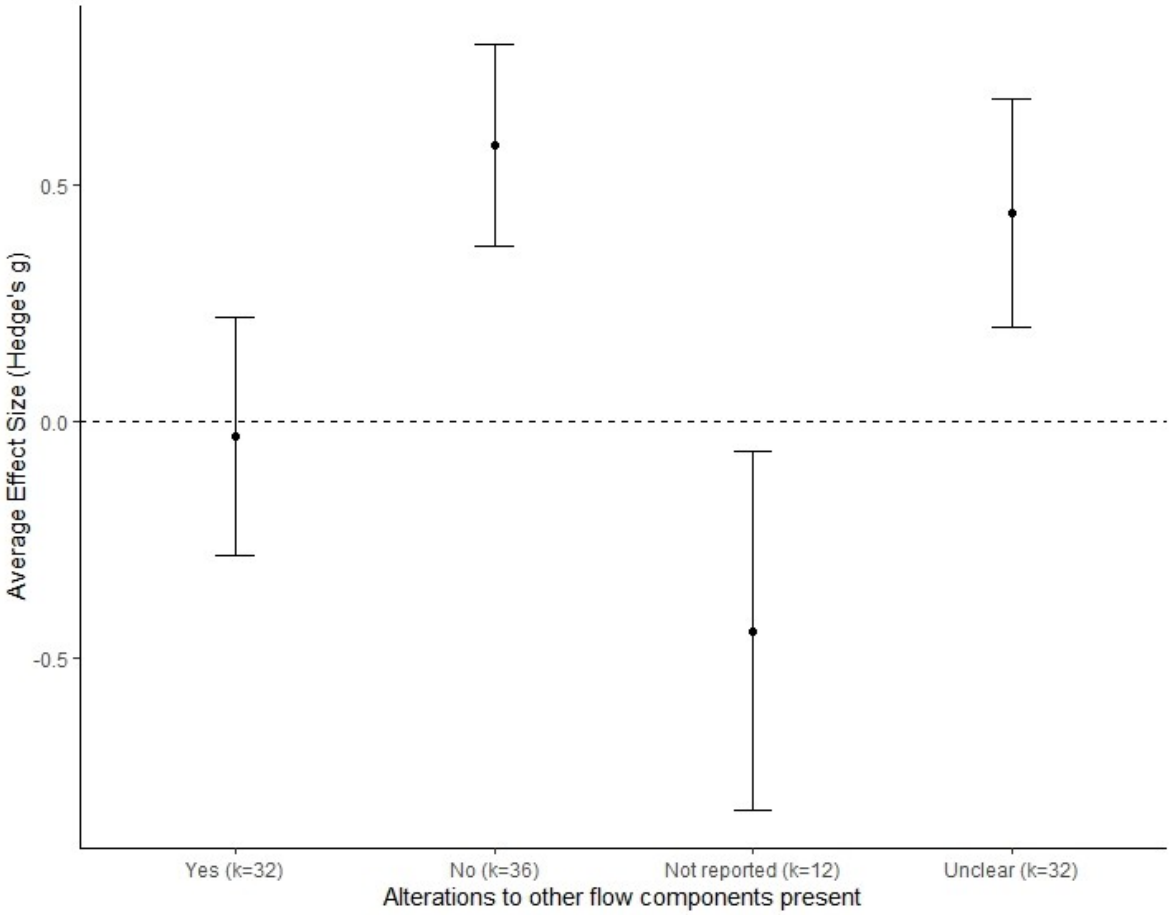

Fig. 17. Average effect size by the presence alterations to other flow components for fish abundance responses when considering interannual *BA* studies. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (above the dashed zero line) indicates that the abundance was higher in the *After* period (intervention) than in the *Before* period (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level). *Yes*: alterations to other flow components were present; *No*: no alterations to other flow components were present; *Not reported*: no mention of alteration to other flow components; *Unclear*: unclear whether another flow component was altered but other flow components are mentioned by authors.

Sampling methods

There were only sufficient sampling sizes and variation to include following levels within sampling methods: (i) gear + angling; (ii) visual; (iii) other (i.e., historical fishing data or hydroacoustics); and (iv) multiple (any combination or two or more methods). Sampling method was associated with average effect sizes (Table 13), but the response of fish abundance to various methods varied (Fig. 21 and Table 13). Visual sampling methods were associated with the largest positive effect sizes, indicating that when using visual sampling methods average fish abundances were larger after the intervention than before. Similarly, gear + angling was associated with positive average effect sizes (Fig. 18). 'Other' sampling methods was associated with negative average effect sizes, indicating the opposite response in abundance when those methods were utilized (Fig. 18).

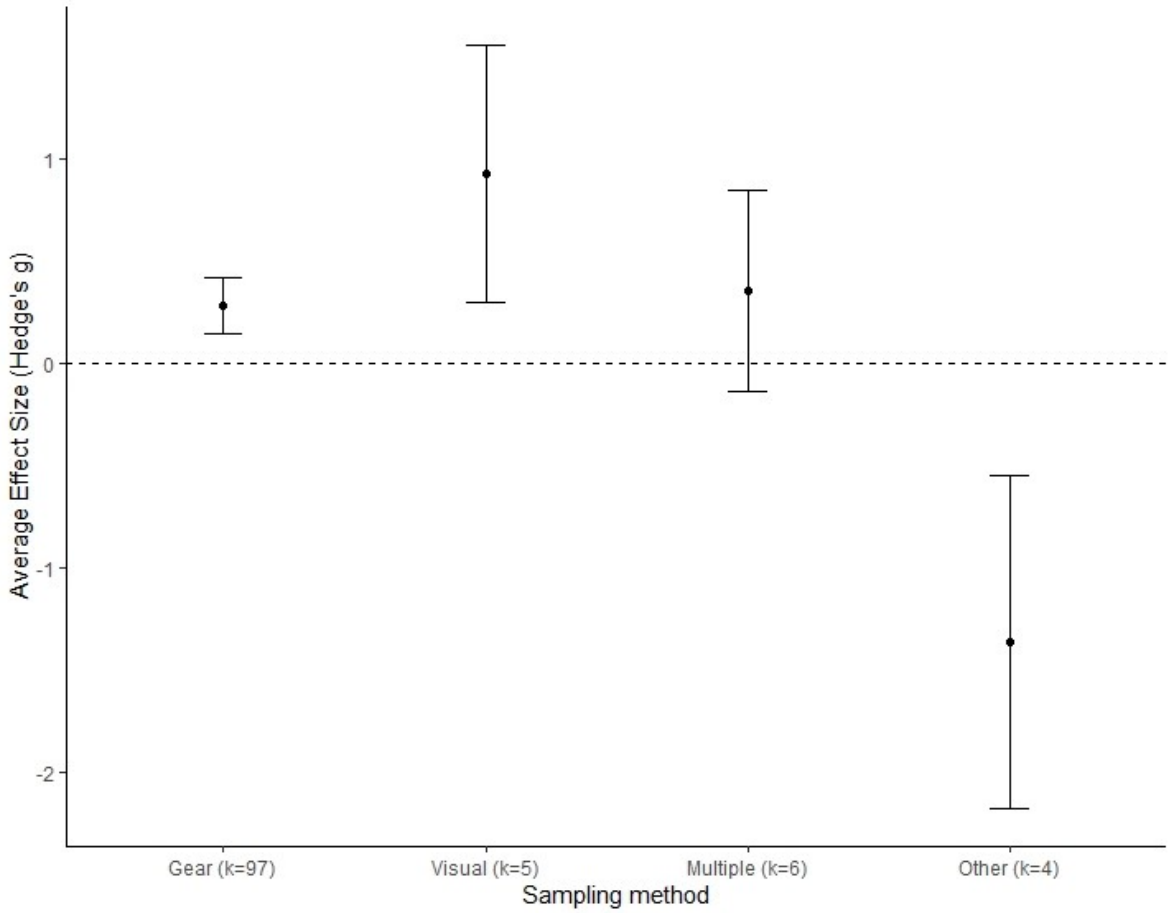

Fig. 18. Average effect size by sampling method for fish abundance responses when considering interannual *BA* studies. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (above the dashed zero line) indicates that the abundance was higher in the *After* period (intervention) than in the *Before* period (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level). *Multiple*: two or more different sampling methods used; *Othe*r: any other methods not previously mentioned (i.e., historical catch data or hydroacoustics).

Sampling seasons

There were only sufficient sample sizes to allow inclusion of the following categorical levels: (i) Spring; (ii) Summer; (iii) Fall; (iv) Multiple (any combination of two or more seasons); and (v) Unclear (not reported + unclear). Sampling season was associated with average effect sizes (Table 13 and Fig. 21). The response of fish abundance to all seasons was positive (i.e., fish abundance was higher in the *After* period than in the *Before* period), with studies conducted in summer having the highest average effect size (other than studies that were not clear regarding

which seasons were sampled) (Fig. 19). Studies conducted in multiple seasons were also associated with statistically significant average effect sizes (potentially because of the inclusion of summer samples within this group), although abundance of fish was lower than when only summer was considered (Fig. 19).

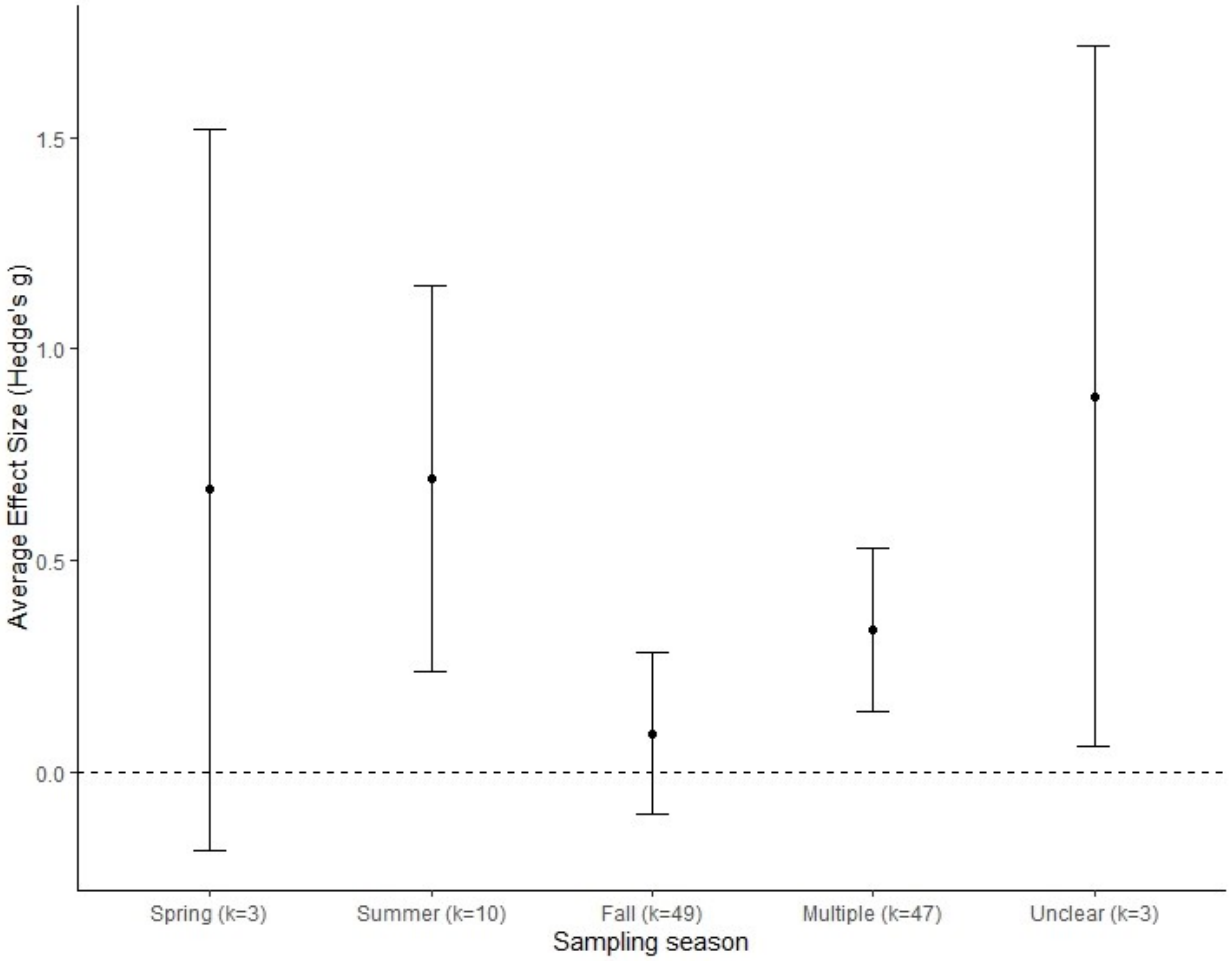

Fig. 19. Average effect size by sampling season for fish abundance responses when considering interannual *BA* studies. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (above the dashed zero line) indicates that the abundance was higher in the *After* period (intervention) than in the *Before* period (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level). *Multiple*: two or more different sampling seasons used; *Unclear*: no clear description of sampling season included.

Type of temporal comparator
We found no detectable effect of type of temporal comparator on average effect sizes (Table 13).

Time since intervention
There were only sufficient sample sizes within time since intervention to consider <1 year, one year, and two years after intervention. We found no detectable effect of this moderator on average effect sizes (Table 13).

Monitoring duration
Due to two extreme outliers in the effect sizes (*Ictalurus punctatus* and *Micropterus dolomieu*; Bestgen et al. 2006), we were unable to achieve normality through transformation for this continuous moderator. We therefore conducted meta-regression with and without these outliers and present results for both analyses in Appendix 11 (see section "Interannual *BA*: Abundance - Meta-regression" and Fig. S28 and S29). We found no significant influence of fish abundance and monitoring duration in either instance and the results of the two models did not differ greatly. We report results of Q_M for the model with outliers in Table 13, but see also Appendix 11 (Table S3).

Life stage
There were only sufficient sample sizes to consider the following life stages: (i) age-0; (ii) larvae; (iii) adult; (iv) mixed (any combination of life stages not reported separately); and (v) not reported (no specific life stage provided). Life stage was associated with average effect sizes (Table 13) with responses in fish abundance varying by life stage (Fig. 20 and 21). Adult fish were associated with larger, positive average effect sizes compared to mixed life stages (which had a moderate positive association with average effect size). Larvae were associated with the largest, most negative average effect size, indicating average fish abundance was lower for larvae after an intervention than prior to a change, but sample sizes were small (Fig. 20).

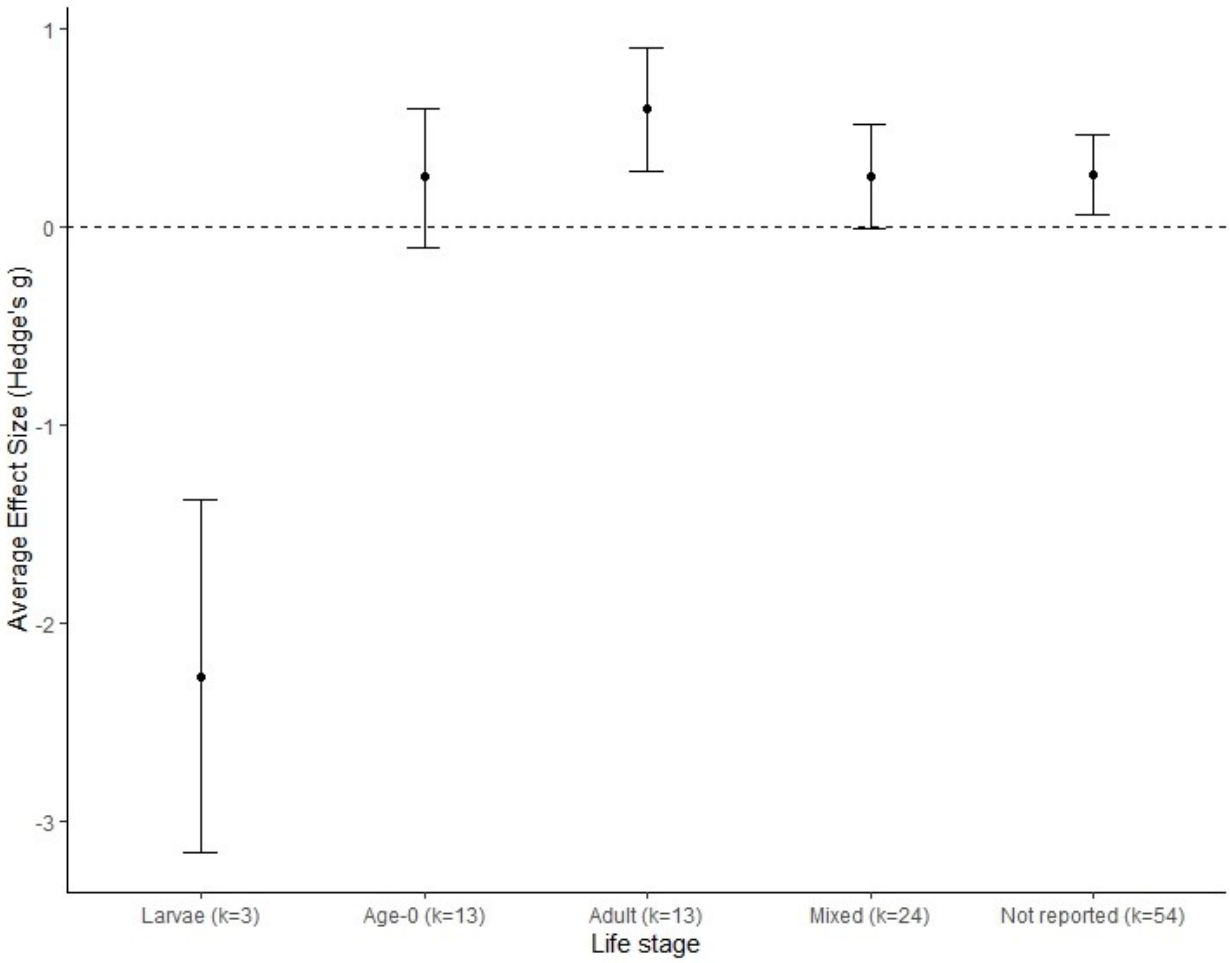

Fig. 20. Average effect size by fish life stage for fish abundance responses when considering interannual *BA* studies. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (above the dashed zero line) indicates that the abundance was higher in the *After* period (intervention) than in the *Before* period (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level). *Mixed*: two or more life stages not reported separately; *Not reported*: no life stage specified

Table 13. Summary results of meta-analysis using subsets of fish abundance effect sizes for interannual *Before/After* studies, testing the influence of the given moderator variable. Bold indicates statistical significance.

Moderator	***k***	***Q* statistic (*p*-value)**	**Q_M (*p*-value)**	**Q_E (*p*-value)**
Waterbody type	112	N/A		
Dam size	112	N/A		
Hydropower operational regime				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Operational regime	112	-	1.35 (*p*=0.717)	**412.00 (*p*<0.0001)**
Direction of flow magnitude alteration				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Flow alteration	112	-	**9.07 (*p*=0.028)**	**412.05 (*p*<0.0001)**
Alterations to other flow components				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Other components	112	-	**29.42 (*p*<0.0001)**	**391.70 (*p*<0.0001)**
Sampling method				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Sampling technique	112	-	**19.83 (*p*=0.0002)**	**401.29 (*p*<0.0001)**
Sampling season				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Sampling season	112	-	**10.11 (*p*=0.039)**	**411.01 (*p*<0.0001)**
Type of comparator (temporal)				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Temporal comparison	112	-	1.63 (*p*=0.203)	**420.92 (*p*<0.0001)**
Time since intervention				
Unmoderated model	111	**420.78 (*p*<0.0001)**	-	-
Time since intervention	111	-	1.96 (*p*=0.376)	**418.83 (*p*<0.0001)**
Monitoring duration (*with outliers*†)				
Unmoderated model	112	**421.12 (*p*<0.0001)**	-	-
Monitoring duration	112	-	1.32 (*p*=0.252)	**417.43 (*p*<0.0001)**
Life stage				
Unmoderated model	108	**415.73 (*p*<0.0001)**	-	**-**
Life stage	108	-	**35.36 (*p*<0.0001)**	**380.12 (*p*<0.0001)**

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. Significance at $p < 0.05$; * Significance at $p < 0.1$. † two extreme effect sizes were removed to improve model fit, but removal had little impact on results.

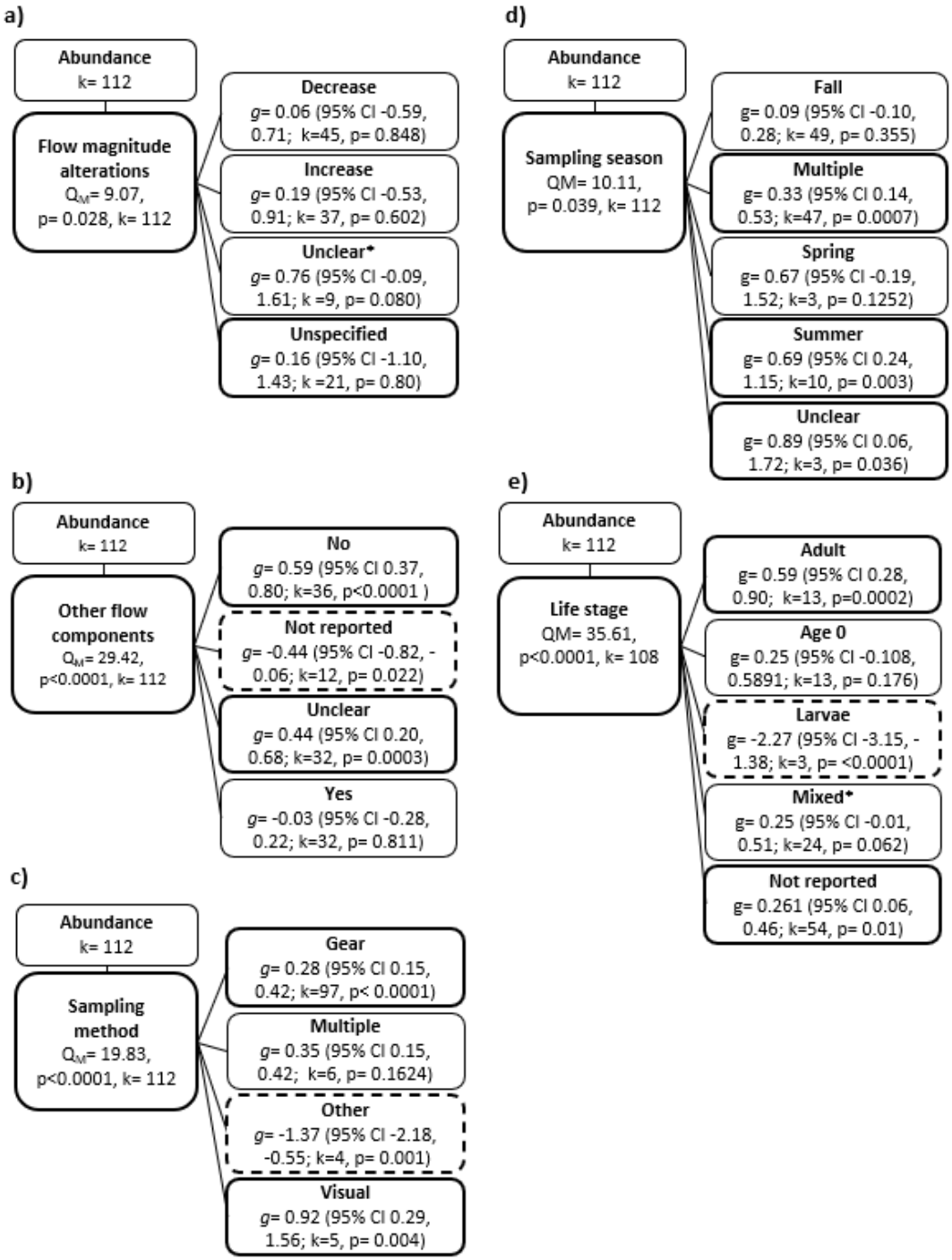

Fig. 21. Summary flow chart of univariate mixed models and resulting significant moderators: (a) direction of flow magnitude alterations; (b) presence of alterations to other flow components; (c) sampling method; (d) sampling season; (e) life stage. *Indicates moderately significant effect (p<0.1). Dashed boxes indicate statistically significant negative effect, thick solid line boxes indicate statistically significant positive effect (i.e., fish abundance is greater in the *After* period than the *Before* period). *k*: number of datasets (i.e., effect sizes); *g*: Hedges' *g* mean effect size; CI: 95% confidence interval.

5.3.6 Taxonomic analysis

Forest plots for all analyses are presented in Appendix 12.

<u>*Control/Impact studies*</u>

There were only sufficient sample sizes to investigate impacts of alterations to flow magnitude due to HPP facilities on abundance for five temperate freshwater fish families for *CI* studies: (i) Catostomidae, (ii) Centrarchidae, (iii) Cyprinidae, (iv) Percidae, and (v) Salmonidae. The families Catostomidae and Cyprinidae had overall negative responses, while Percidae, Centrarchidae and Salmonidae families had overall positive responses to flow magnitude alterations, although no family had a statistically significant overall response (Table 7 and Fig. 22). Based on the Q test of heterogeneity, heterogeneity among effect sizes for the Catostomidae family was not significant (Q=3.95, p = 0.266). Three species (*Moxostoma poecilurum*, *Carpiodes velifer*, and *Catostomus occidentalis*) and one unknown species were present in this group. Similarly, there was no statistically significant heterogeneity among effect sizes for species of Centrarchidae family (Q = 1.32, p = 0.93), which was represented by five species from two known genera [*Lepomis* (k= 2) and *Micropterus* (k=3)] and one unknown genera, Percidae family [represented by three known and one unknown genera (Q = 2.2745, p = 0.5174)] or Salmonidae family (Q = 7.33, p = 0.119) which was primarily *Salmo trutta* (5 of 6 datasets). In contrast, the Q test for heterogeneity indicated that there was significant heterogeneity among species effect sizes for Cyprinidae (Q = 39.41, p = 0.003) which could be explored through moderator analysis, however 14 of 19 effect sizes were from a single study and there was insufficient variation in moderators to allow effective evaluation of the influence of moderator variables within the outcome subgroup. Sample sizes were too small for moderator analysis of other families and abundance, or for analyzing biomass responses by taxa.

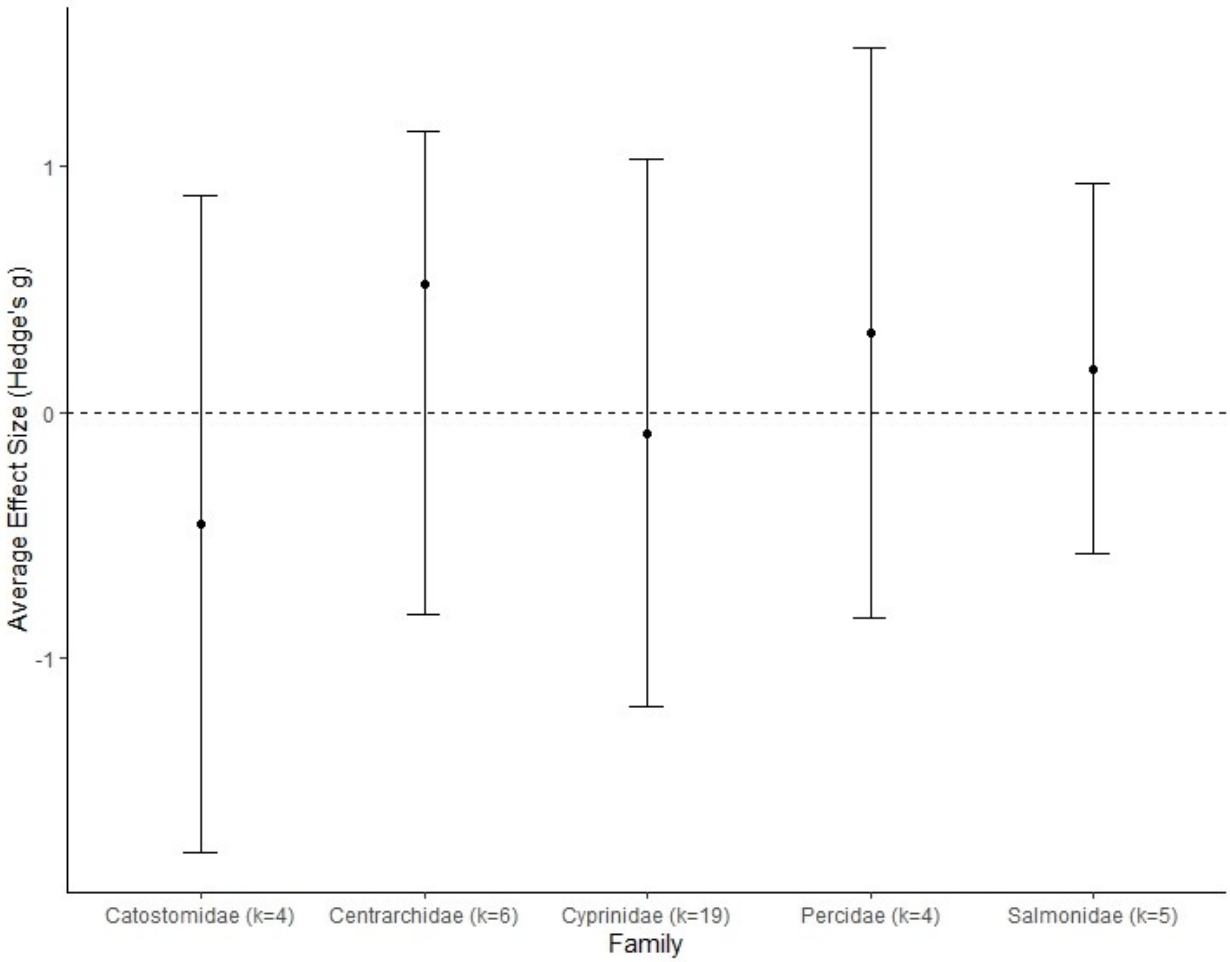

Fig. 22. Average effect size by fish family for *Control/Impact* studies and abundance. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (above the dashed zero line) indicates that the abundance was higher in intervention than in comparator sites (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level).

<u>*Within-year Before/After studies*</u>

There was only sufficient sample size to investigate impacts of alterations to flow magnitude due to HPP facilities on abundance responses for one temperate freshwater fish family, Salmonidae, for *BA* studies. Salmonidae had a moderately significant negative response to flow magnitude alterations [Hedge's $g = 0.81$ (95% CI -0.15, 1.76; $k = 4$, $p = 0.099$)] (Table 7). This may be due to a single statistically significant positive effect size related to *Oncorhynchus mykiss* in the Colorado River, where this species is an established non-native species, previously stocked, but

now established (Avery et al. 2015) (Appendix 13: Fig. S6). Based on the Q test of heterogeneity, heterogeneity within this family was not statistically significant. This may be due to the presence of only two species in this group of datasets: (i) *Oncorhynchus mykiss* (k=3) and (ii) *Salmo trutta* (k=1). Sample sizes were too small in this group to evaluate influences of moderator variables within the abundance outcome for this family. No within-year *BA* studies considered biomass outcomes.

Interannual Before/After studies

Abundance

There were only sufficient sample sizes to investigate impacts of alterations to flow magnitude due to HPP facilities on abundance for nine temperate freshwater fish families for interannual *BA* studies: (i) Acipenseridae; (ii) Anguillidae; (iii) Catostomidae; (iv) Centrarchidae; (v) Cottidae; (vi) Cyprinidae; (vii) Esocidae; (viii) Ictaluridae; and (ix) Salmonidae. For families with significant heterogeneity among effect sizes (i.e., significant Q), additional analyses were performed for genera therein with sufficient sample size (i.e., ≥3 datasets from ≥2 independent studies) (Appendix 12: Fig. S17 – S22).

The families Acipenseridae, Centrarchidae, Esocidae and Ictaluridae had overall positive but nonsignificant responses to alterations in flow magnitude (i.e., fish abundance was greater after an intervention than prior to the intervention) (Table 7 and Fig. 23). Centrarchidae [Hedges' g = 7.14 (95% CI -5.68, 19.96; k = 5, p = 0.275) and Ictaluridae [Hedges' g = 6.69 (95% CI -6.86, 20.24; k = 3, p = 0.333)] had strong positive responses to flow alterations, but the heterogeneity for both was significant and much larger than for the other families considered (Centrarchidae: Q = 29.06, p <0.0001; Ictaluridae: Q = 30.34; p <0.0001) (Fig. 23). Anguillidae, Catostomidae and Cyprinidae all had negative overall mean effect sizes, but also had

nonsignificant responses to alterations in flow magnitude (Table 7 and Fig. 23). In contrast, the overall weighted mean effect size for Salmonidae was 0.45 (95% CI 0.25, 0.65; $k = 59$, $p < 0.0001$), suggesting that alterations to flow magnitude had a positive and significant effect on salmonid abundance (i.e., abundance was greater after an intervention than it was prior to the intervention).

Based on the Q test of heterogeneity, there was also significant heterogeneity among effect sizes for Acipenseridae ($Q = 31.09$, $p < 0.0001$) which included only two species (*Acipenser fulvescens*, *Acipenser sinensis*), Catostomidae ($Q = 70.26$, $p < 0.0001$) which included five species from one genus, Cyprinidae [15 species from 11 genera due to combined species outcomes; ($Q = 87.3019$; $p < 0.0001$)], and Salmonidae [13 species from five genera ($Q = 108.13$; $p < 0.0001$)]. Although heterogeneity was present within these families, we only conducted analyses at the genera level when more than one genus with sufficient sample sizes were present (i.e., Cyprinidae and Salmonidae; Appendix 12 section 'Interannual *Before/After*: Genera'). Anguillidae (one species; $Q = 5.88$, $p = 0.21$), Cottidae (one genus; $Q = 8.15$, $p = 0.086$) and Esocidae (one species; $Q = 0.76$; $p = 0.685$) did not have statistically significant heterogeneity (Fig. 23).

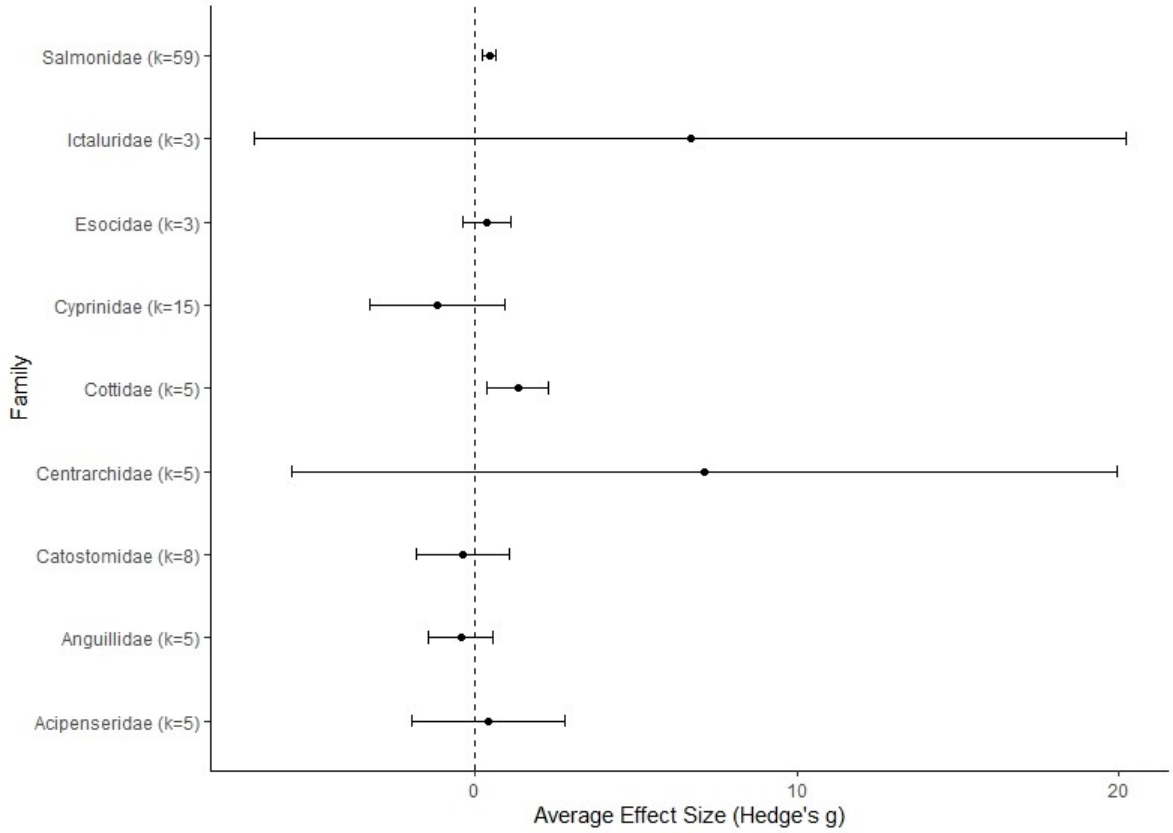

Fig. 23. Average effect size by fish family for interannual *Before/After* studies and abundance. Value in parentheses (*k*) is the number of effect sizes. Error bars indicate 95% confidence intervals. A positive mean value (right of the dashed zero line) indicates that the abundance was higher in the *After* period (intervention) than in the *Before* period (no intervention). 95% confidence intervals that do not overlap with the dashed line indicate a significant effect (at the $p>0.05$ level).

Biomass

There was only sufficient sample size to investigate impacts of alterations to flow magnitude due to HPP facilities on biomass responses for one temperate freshwater fish family, Salmonidae, for interannual *BA* studies. The overall weighted effect size for Salmonidae was 0.52 (95% CI -0.38, 1.43; $k = 11$, $p = 0.258$) (Table 7), which indicates that alterations in flow magnitude may have a positive effect on fish biomass, but that the response was not significant. There was no statistically significant heterogeneity among effect sizes based on the *Q* test of heterogeneity (*Q*

= 15.02; p = 0.131). There was sufficient sample size to investigate one genus in this family (*Oncorhynchus*; Appendix 13: Fig. S23).

6. Discussion

Our comprehensive analysis of the effect of flow magnitude alterations due to hydropower production and operations on fish abundance and biomass did not find consistent patterns in how fish respond to increases and decreases in flow magnitude. Although the impact of flow alterations on fish responses has been previously reviewed (see Poff and Zimmerman 2010; Young et al. 2011; McManamay et al. 2013; Webb et al. 2013), our approach differed in that it focused on a few specific responses of fish (i.e., abundance and biomass) and on a single flow component (magnitude), allowing us to target this one aspect of the fish-flow relationship. Additionally, our approach improved on past reviews by providing an extensive, systematic search of the available literature, using a rigorous, objective and transparent methodology, resulting in a more comprehensive analysis of these specific dimensions of the flow alteration-ecological response relationship. Moreover, individual studies were subjected to critical appraisal to determine their reliability and, as necessary, biased studies were down-weighted.

We identified 133 relevant studies, of which 58 were eligible for quantitative analysis. We acknowledge that our review does not represent the entirety of the knowledge base on the subject; for example, we excluded studies with qualitative indicators of changes in fish responses (i.e., presence/absence of fish) from our quantitative synthesis. Nonetheless, those studies did inform our narrative review. To ensure that our review identified and included the most relevant and reliable studies (i.e., studies with low internal bias) to answer our specific question, during the screening process we also excluded studies without spatial or temporal comparators (Fig. 2).

These studies may have useful information on the topic and are an important part of the existing literature, but did not meet our stringent requirements.

6.1 Impact of flow magnitude alteration on fish responses

Fish responses to alterations in flow magnitude were variable (Table 7). Overall mean effect sizes ranged from positive to negative and varied depending on the outcome (abundance or biomass), the study design (*CI*, within-year *BA* or interannual *BA*) and the taxa considered (Table 7). For *CI* studies, alterations in flow magnitude led to effect sizes that indicated almost no change or negative changes in abundance, while fish biomass was estimated to have a negative overall effect size (i.e., was lower relative to comparators not receiving an intervention). In contrast, for *BA* study designs, both abundance and biomass had generally higher values in the *After* period relative to the *Before* period for within-year and interannual *BA* study designs (i.e., a positive overall response to flow magnitude alterations). This difference may arise because *CI* study designs compared fish outcomes at an impacted site to a non-impacted comparator (i.e., upstream of the HPP facility or a different unimpacted waterbody), whereas many BA studies reported alterations to flow at existing HPP facilities that were made specifically to provide potential benefits to fish (i.e., increases in base flow). None of the overall effect sizes from *CI* or interannual *BA* studies were statistically significant ($p<0.05$), but abundance in *After* year-1 and *After* year-2 of within-year *BA* studies did have moderately significant effect sizes, although sample sizes were small (Table 7). These results are consistent with past reviews, which also found that fish responses to flow magnitude alterations were highly variable and context dependent (McManamay et al. 2013; Webb et al. 2013; Turgeon et al. 2019). It has been argued that over-generalization or simplified 'rules-of-thumb' applied across systems should be avoided (Arthington et al. 2006) and, because river systems have unique physical properties (Schumm

2005; Konrad et al. 2011) and communities (Oberdorff et al. 2011; Nicol et al. 2017), knowledge of one system cannot necessarily be transferred to other systems.

Taxonomic responses varied across families, although interestingly, responses of specific taxa were consistent across *CI* and interannual *BA* studies (Table 7). Overall mean effect sizes for Catostomidae indicated a decrease in abundance relative to a comparator in both *CI* and *BA* study designs, as did those for Cyprinidae, while effect sizes for both Centrarchidae and Salmonidae saw overall increases relative to comparators (Table 7). Caution should be used when interpreting effect sizes for most taxonomic groups however, as sample sizes were small. The overall responses of these families may indicate that, although a generalizable trend across taxa may not be possible, specific families or species may respond consistently to changes in flow magnitude. For example, low heterogeneity within the family Salmonidae (Fig. 23) may indicate that salmonid abundance responds positively to alterations in flow magnitude. However, caution should be taken when interpreting this result, because we were unable to explore potential moderators for this subgroup and sample sizes were small.

Several moderators were tested in our quantitative synthesis to explore reasons for heterogeneity among responses. Moderator analysis for *CI* studies was inconclusive, with no detectable effect of any moderators concerned with dam operations (i.e., dam size, hydropower operational regime, direction of flow magnitude change), potential confounders (i.e., alterations to other flow components, time since sampling) or study design/methods (i.e., sampling season or method, type of comparators used, or monitoring duration). In contrast, several moderators were associated with interannual *BA*s. Five moderators were associated with the overall effect sizes for abundance: (i) direction of flow magnitude alterations; (ii) presence of alterations to other flow components; (iii) sampling method; (iv) sampling season; and (v) life stage. However,

considerable residual heterogeneity remained in the observed effects of hydropower production, suggesting that interactions between moderators may be occurring or that some other factors, not captured in our analysis, may be influencing fish abundance. Additionally, most moderators were highly correlated (see Pearson's χ^2 test Appendix 11: Table S1 and S2), complicating interpretation and making models with multiple variables impracticable due to small sample sizes (Lipsey 2003). It was, therefore, difficult to determine the impact of each moderator on the overall mean effect sizes.

6.2 Potential biases

We attempted to limit potential biases throughout the systematic review process. There was no apparent evidence of publication bias for fish biomass (Appendix 11: Fig S7 and Fig. S25), however sample sizes were small. There was possible evidence of publication bias for abundance in both *CI* and interannual *BA* studies towards studies showing increased abundance in the intervention site relative to controls (Appendix 11: Fig. S2 and S17). When separating publication bias by publication type, evidence of publication bias towards positive results was present in grey literature, but not in commercially published literature. A possible explanation may be that these reports are commissioned by hydropower operators to quantify impacts of flow alterations at their facilities, which may have led to lower reporting of negative results, due to the types of questions being investigated (e.g., practitioners focus on flow improvements), whereas commercially published literature may focus on the overall impacts of flow alterations. It is almost certain that additional grey literature exists (especially for earlier decades where grey literature made up less than 50% of studies identified; Fig. 4) in internal documents that were not accessible to our review team.

To limit availability bias, efforts were made to obtain all relevant materials, however there were several reports and publications (n= 17) that were unobtainable (Appendix 4). Our review was limited to only English articles. We feel that we have captured what is available and relevant to the Canadian and North American context of this review. However, there may be valuable articles and grey literature from other countries in temperate regions that were not published in English. Although we did not use non-English search terms in either the systematic map (Rytwinski et al. 2020) or this review, there were relatively few articles excluded from the map on language at full text (61/2412 articles) and only four excluded on language during this review (Fig. 2). This low number of excluded, non-English studies may suggest that there is only a slight risk of language bias.

A potential seasonal bias was also present in our quantitative synthesis. A number of studies (30/58) considered only a single sampling season (corresponding to 117/256 datasets). Because fish populations change with seasons, focusing on a single sampling period may overemphasize responses that are due to population behaviour or other seasonally influenced environmental factors (Pope and Willis 1996). Additionally, several studies in our quantitative database used comparators which were sampled in different seasons for the same species (5 studies, 19 datasets). This may lead to additional issues in interpreting responses if a species goes through episodic population fluctuations or variable seasonal reproduction (Barrett and Munkittrick 2010). Long term, multi-seasonal studies could help alleviate these risks and studies conducted in a consistent season over a longer period of time can provide useful insight into general population changes (e.g., pink salmon; Morsell 2000). Winter fish sampling was underrepresented in both the narrative synthesis (11% of cases) and quantitative synthesis (11% of datasets), and was rarely reported individually to isolate fish responses during this season.

There is a general lack of knowledge on the importance of winter in fish population dynamics (Block et al. 2020) and fish responses to flow alterations during this period were comparatively missing in our database (although see Reyjol et al. 2001). The lack of sampling during this period, potentially due to logistical and methodological challenges associated with sampling fish during winter (Block et al. 2019; Studd et al. In press), may bias apparent fish responses to flow alterations and limit our understanding of how altered flow regimes impact fish. A recent systematic review on the seasonal phenology of fish (Brady et al. 2020) indicated a similar bias towards summer sampling and against winter sampling as was identified in ours, which may ultimately limit our ability to fully understand species responses to complex flow regimes.

Geographical and taxonomic biases were evident in the quantitative synthesis. Datasets were primarily from North America (71%) of which 58% were from the United States. Additionally, a single dataset considered both river and estuary sites. Although 98 species were represented within the quantitative synthesis, only six species had more than eight datasets, five of which were from the Salmonidae family (30%). Similar geographic and taxonomic biases were also identified in the systematic map (Rytwinski et al. 2020) and have been identified in other hydropower related reviews for temperate regions (e.g., Algera et al. 2020). Again, this may limit the applicability of our review results for other geographic regions and taxa.

6.3 Limitations of review and evidence base

Collectively, the studies reviewed here did not provide clear insight into the impact the direction of flow magnitude alterations would have on fish and whether the apparent increase in fish abundance (Fig. 16) was actually attributable to flow alterations, or to some other factor such as sampling season. One factor that may have influenced the results is the relatively low validity of included studies. Of the datasets included for quantitative analysis 51% had medium validity,

while the remainder had low validity. When medium validity studies are considered alone, the relative magnitude of the effect size for abundance in *CI* studies is higher, when compared to the overall model (i.e., an increase from g= -0.001 to g= -0.18 when low validity studies are removed; Table 8). This indicates that the inclusion of low validity studies may lead to smaller effect sizes overall, although the overall effect sizes were non-significant in either case. Improving study design by including both temporal and spatial replication, improving comparator matching and improving reporting of flow alterations would all aid in improving internal validity of primary studies and, subsequently, increase the reliability of future evidence syntheses.

Our inability to clearly identify fish-flow relationships may, in part, be due to our inability to quantify the amount of flow alteration experienced (i.e., ΔQ, where Q is discharge). However, we felt that our qualitative descriptions of flow alteration based on author descriptions allowed us to capture a larger percentage of studies for quantitative analysis than would otherwise have been possible, and the inclusion of both qualitative and quantitative flow descriptions is not uncommon in the review literature on this topic (e.g., flow categories; McManamay et al. 2013). Effectively quantifying differences in flow magnitude was complicated by inconsistent reporting of flow alterations among studies. For example, several studies quantified flow magnitude or included hydrographs for only the *After* period, or the intervention sites, without also including some measure of flow for the comparator. This made calculations of the amount of flow alteration impossible. In other instances, historical hydrographs were included, but only qualitative descriptions of the specific flow alteration being investigated were reported. We recommend that, when reporting alterations in flow magnitude, comparable data (i.e., measured flow magnitude or hydrographs) from the same temporal period

(i.e., season) be included for both the intervention and comparator sites or periods. Expanding and improving flow monitoring systems throughout impacted and unimpacted waterbodies would assist in these efforts.

Many articles were excluded due to choice of comparator or were assessed to have low study validity during critical appraisal because of imperfect matching. Of particular note, 144 articles were excluded during screening due lack of a useable comparator and many due to a complete lack of comparator (i.e., spatial trends or temporal trends). While we acknowledge that much of the literature represented by these studies did not set out to answer questions similar to this review (i.e., how flow alterations from hydropower impact fish outcomes), in cases where a similar question was being asked, and only trends were being considered, important aspects of flow and fish dynamics may have been missed. For example, if alterations in fish abundance downstream are apparent after a flow alteration, but no sampling occurs upstream, a general decline in fish abundance along the entirety of the river, due to other basin-scale factors may be missed. Additionally, several articles were excluded due to the use of downstream comparators. We recommend that researchers limit the use of downstream comparators when studying flow alterations. Although upstream impacts attenuate downstream (Konrad et al. 2011), it is unlikely that downstream sites will ever truly be unimpacted. Indeed, in some cases the impacts of upstream dams are detectable hundreds of kilometers downstream (Williams and Wolman 1987). Instead, upstream or unimpacted comparators are recommended, although they present their own challenges (e.g., site matching or availability).

A challenge in assessing flow alterations and fish responses is knowing if the response seen in fish outcomes can actually be attributed to the alteration in flow (Konrad et al. 2011). If studies focus on spatial replication, changes in responses through time may be missed, while

studies focused on temporal replication may miss underlying spatial variability or change. To properly assess the impacts of flow alterations, including both spatial and temporal replication within the same study is essential (i.e., *BACI* designs). In simulation, *BACI* designs outperform other study designs, including *BA* and *CI* designs, with higher accuracy and less bias (Christie et al. 2019). In an ecological context *BACI* designs have benefits over *BA* or *CI* designs because they help decrease the likelihood of erroneous conclusions based on the inherent assumptions of similarity of spatial sites or *Before/After* periods in these designs (Smokorowski and Randall 2017). We had opposing overall effect sizes when considering studies with different replication (i.e., negative and positive effect sizes for *CI* and *BA* studies, respectively). This may be a function of *BA* study designs potentially providing more accurate results than *CI* designs (Christie et al. 2019) and indicates that for a more complete picture, including both spatial and temporal replication would be helpful to truly understand outcomes. *BACI* studies were less represented in our quantitative synthesis than other study designs (13 cases; see "Study design and Comparator"), likely due to the time and complexities required in these study designs. This limited our ability to examine these studies specifically, or at the interaction level, because they had to be converted for inclusion in analysis.

Most of the *Control/Impact* studies included in this review lacked true spatial replicates and all *BA* studies lacked replication across waterbodies. Pseudoreplication was more common than true replication in the surveyed literature (for one exceptional example of true replication with >30 replicates in both the comparator and intervention see Göthe et al. 2019), but considering only studies with true replication did lead to a larger, overall negative effect size, indicating that including pseudoreplicated studies may result in a smaller overall effect size (Table 8). Although the inclusion of pseudoreplication may lead to issues of nonindependence of

samples (Hurlbert 1984), it may not be possible to identify similar dams with similar operations and sizes in similar hydrological settings and sample from these true replicates during similar periods of dam operation (Konrad et al. 2011). Given that true replication may not be possible in many situations, efforts should be made to at least sample in multiple locations downstream and upstream of the intervention both before and after an intervention (i.e., closure of a new dam or change in existing flow regime) occurs, and to control for pseudoreplication during analysis (Schank and Koehnle 2009). Caution in selecting upstream comparators should be taken, as dams may act as barriers to dispersal and movement (Fuller et al. 2015), and any apparent increase in fish abundance downstream may be due to pooling below the dam and loss of fish upstream, rather than an actual increase in population. In systems where multiple dams are present and their impacts may interact (e.g., Tallapoosa River) or there are dam cascades (e.g., Yangtze River), ensuring that spatial comparators are selected outside the influence of any dam can be extremely difficult, but not impossible (e.g., Bowen et al. 1998). Spatial replication should still be attempted and the potential impacts of upstream facilities should be explicitly stated in any study within these types of systems.

Due to limitations in the existing evidence base, we were unable to draw clear conclusions of time lags or the long-term effect of alterations in flow magnitude. The majority of studies in this review did not report within-year fish responses (i.e., data for each month or season of sampling reported separately for each sampling year), instead providing single data points, sums or averages across within-year sampling without also providing within-year fish outcomes over several years [although see Avery et al. (2015) and Crisp et al. (1983) for examples of within-year replication reported for >3 years]. Additionally, many *CI* and interannual *BA* studies included in quantitative synthesis were based on short-term monitoring

[but see Sullivan and Hileman (2018) for an exception with > 15 years of sampling both pre- and post- intervention]. Long-term studies are important to identify changes in responses through time and help elucidate patterns and other important factors that would otherwise be missing in short-term data (Lohner and Dixon 2013; Counihan et al. 2018). For example, if fish respond differently after several years of exposure to an intervention [as was identified in Ugedal et al. (2008) who saw immediate declines in fish density post-intervention followed by a gradual recovery over a period of nearly 20 years], short-term studies may not capture these changes. This is especially true if a single age group or single sampling period is considered (Lohner and Dixon 2013). Decreases in population may not become immediately apparent in long-lived species if only adults of the species are considered for a short, 1-2 year period; conversely, potential benefits of alterations (such as increases in base flow) may increase in value over time as fish become adapted to a new flow regime. Of all studies included in quantitative analysis, a single study reported >4 years (Crisp et al. 1983) that could be used to assess time lags, and only two *CI* studies reported ≥3 years. Interannual *BA* studies were often longer in duration, with six studies lasting a decade or more. These types of studies should continue to be encouraged. When paired with flow experiments, such as those conducted in the Colorado River system [see Korman et al. (2011), Avery et al. (2015) and associated supplemental material; Appendix 5], these types of studies can expose aspects of responses that would otherwise be obscured and even open new avenues of research (Hampton et al. 2019).

6.4 Further considerations

6.4.1. Composition

Other fish responses may be occurring that are not apparent in the outcomes we focused on here. If individual fish species or taxa respond differently to flow alterations, as seen in our taxonomic analyses for *CI* and *BA* studies, compositional changes in fish populations may result. Changes in

composition have been seen in impoundments (Turgeon et al. 2019) and assemblage dominance and species composition has been found to differ between sites upstream and downstream of HPP facilities, or after flow alterations (Travnichek et al. 1995; Enders et al. 2017). Recent estimates of biodiversity changes in freshwater systems due to human-induced alterations indicate that temperate regions have experienced among the largest biodiversity changes of any region (Su et al. 2021). How different populations are responding to hydropower production and operations in terms of compositional changes is an area for further exploration.

6.4.2 Recommended study design

To better assess the impacts of flow magnitude alterations on fish outcomes through systematic reviewing and quantitative synthesis, we make the following recommendations for future study design and reporting:

Controls – Authors should make every effort to incorporate temporal and/or spatial comparators in their studies to ensure adequate baselines and improve understanding of impacts. Although difficult and resource intensive, full *BACI* study designs are essential to properly account for temporal and spatial confounders. If not possible, selecting more accurate study designs [e.g., *BA* are considered more accurate than *CI* studies (Christie et al. 2019)] with comparators that are carefully matched will facilitate more accurate quantitative synthesis results. Care should be taken to avoid downstream comparators whenever possible and to minimize gaps between temporal sampling periods and interventions.

Duration – When designing studies to assess fish responses to flow magnitude alterations, long-term monitoring (i.e., >2 years) both prior to and post-intervention would facilitate improved understanding of population level effects and time-lags in responses. This is especially important

for longer-lived species. Efforts should be made to minimize gaps between sampling years, and to ensure sampling occurs in multiple seasons.

Replication – Care should be taken to ensure that appropriate levels of replication are included. Authors should ensure replication occurs in both the intervention and comparator, to facilitate inclusion of more studies in quantitative synthesis. When combining studies in syntheses, more accurate results (i.e., those from true replication) are preferable to more precise results (i.e., those from pseudoreplication). However, as true replication is not always possible, authors who find themselves in situations where true replication is unobtainable should still aim to include replicate sampling (even if pseudoreplicates).

Reporting – Studies should report sufficient detail regarding location of sample sites (i.e., latitude and longitude), the degree of replication (true or pseudoreplicates) and, when possible, report summarized data separately for monthly or seasonal samples within a year, and report detailed descriptions of how samples are grouped for analysis or report raw data. Authors should make every effort to include a detailed description of all hydropower facility design and operations as well as both qualitative and quantitative descriptions of flow alterations. Studies must as adequately describe the pre-intervention and comparator site as they do the post-intervention and intervention sites, in order to capture vital information on confounding factors or differences in starting conditions. Where information cannot fit within published articles, details should be included in supplementary materials.

6.5 Implications and conclusion

Systematic reviews with meta-analysis aim to generalize ecological relationships and explore differences in individual study characteristics and heterogeneity in results (CEE 2018) but in some instances, results of systematic reviews may be ambiguous (Rytwinski et al. 2021) and

generalizations may not be possible. Previous reviews on fish/flow relationships determined that generalizable and transferable relationships between flow components and species responses were not possible with the state of the literature base (Poff and Zimmerman 2010) and that relationships were highly context dependent (McManamay et al. 2013). Nearly a decade later, and with a more extensive, targeted review considering a single flow component, the results of our review are consistent with these findings. Generalizable signals were very difficult to identify and generalization may not be possible in systems impacted by hydropower facilities where the specific features of the system (i.e., size, underlying hydrology, community dynamics) are highly influential. To compensate, regional, long term, continuous monitoring to inform decision making will help improve clarity. Adaptive management and long-term experimental flow studies can further aid decision makers in learning more about their specific systems (Hampton et al. 2019) and in developing flows that provide for both energy and ecological needs (Acreman et al. 2014). Work should continue to grow the evidence base of fish/flow relationships, but should focus on long-term, high quality site and species specific efforts to improve our understanding of how specific species in specific locations interact with flow. Our inability to identify generalizable trends, even with our comprehensive approach, lends credence to the need for sustained, high quality regional science for supporting management decisions in systems impacted by flow magnitude alterations due to hydropower production and facility operation. Although it would be desirable to identify general science-based ecological rules and relationships that extend across regions and taxa, it is evident that such goals remain elusive in the context of fish-hydropower interactions.

7. Availability of data and materials

Appendices 3, 6, 7, 9, 11 and 12 are included immediately below.

Appendices 1, 2, 4, 5, 8 and 10 with data supporting the results of this systematic review are available from OSF: https://osf.io/g3cs8/?view_only=96edd7514068451eaadba5a1bf9d8f54

We provide descriptions of all appendices and their content below.

8. Ethics approval and consent to participate

Not applicable

Appendix 1. ROSES systematic review checklist

Description: This appendix provides the ROSES systematic review checklist (Haddaway et al. 2017) for this review. See OSF link in section "Availability of data and materials".

Appendix 2. Systematic review protocol

Description: Published protocol for the systematic review. See OSF link in section "Availability of data and materials".

Also available from:

https://environmentalevidencejournal.biomedcentral.com/articles/10.1186/s13750-020-00198-5

Citation:

Harper, M., T. Rytwinski, J. J. Taylor, J. R. Bennett, K. E. Smokorowski, and S. J. Cooke. 2020. How do changes in flow magnitude due to hydroelectric power production affect fish abundance and diversity in temperate regions? A systematic review protocol. Environmental Evidence 9(1):14.

Appendix 3. Search strategy and results

Description: This appendix provides a description of the search strategy and results of the literature searches. For each source, we provided full details on the search date(s), search strings used, search settings and restrictions, and subscriptions (if applicable), and the number of returns. We also include the original search string used in the systematic map (Rytwinski et al. 2017).

Databases

The search string was developed based on suggestions from the Advisory Team as described in the protocol of the review (Harper et al. 2020).

The following bibliographic databases were searched from October - December 2019 using Carleton University's institutional subscription

1. ISI Web of Science Core Collection—multidisciplinary research topics including journals, books, proceedings, published data sets and patents
2. ProQuest Dissertation & Theses Global—collection of dissertations and theses from around the world, spanning from 1743 to present
3. Scopus—abstract and citation database of peer-reviewed literature including journals, books, and conference proceedings.
4. Federal Science Library (Canada)—Canadian government books, reports, government documents, theses, conference proceedings, and journal titles
5. Science.gov—US Federal Science
6. AGRICOLA (Agricultural Research Database)—US Department of Agriculture's National Agricultural Library

Web of Science: Search Strategy #1

Note: "Topic" search in Web of Science includes: title, abstract, keywords, keywords plus

Table S1. Metadata from Web of Science Search Strategy #1

Search string	Restrictions	Returns [Date]

TS=((Fish*) AND ("Fresh water" OR Freshwater OR Stream$ OR Water$ OR River$ OR Fluvial OR Estuar* OR Reservoir$ OR Impoundment$ OR "Hydro electric*" OR Hydroelectric* OR "Hydro dam*" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam$) AND (Flow* OR Discharg*) AND (Productivity OR Biomass OR Abundance$ OR Densit* OR Yield$ OR "Ecological response" OR "Ecosystem response" OR "Biotic response") NOT (mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon))	• 2017-2019
	• Web of Science Core Collection
	• Advanced search
	• Topic field
	• All languages
	• All document types
	• Institution subscriptions:
	  o Science Citation Index Expanded (1900 - present)
	  o Social Sciences Citation Index (1956 - present)
	  o Arts & Humanities Citation Index (1975 - present)
	  o Conference Proceedings Citation Index - Science (1990 - present)
	  o Conference Proceedings Citation Index - Social
		818
		[December 1, 2019]

	Science and Humanities (1990 - present) o Book Citation Index - Science & Social Science (2008 - present) o Current Chemical Reactions (2008 - present) o Index Chemicus (2008 - present) o Part of the larger Web of Science.	

ProQuest Dissertations & Theses Global: Search Strategy #2

Note: Command line advanced search selected to search within: title, abstract, keywords.

Table S2. Metadata from ProQuest Dissertations & Theses Global Search Strategy #2

Search string	Restrictions	Returns [Date]
TI,AB,IF(Fish* AND (“Fresh water” OR Freshwater OR Stream OR Water OR River OR Fluvial OR Estuar* OR Reservoir OR Impoundment OR "Hydro electric*" OR Hydroelectric* OR "Hydro dam*" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam) AND (Flow* OR Discharg*) AND (Productivity OR Biomass OR Abundance OR Densit* OR Yield OR “Ecological response” OR “Ecosystem response” OR “Biotic response”) NOT (mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon))	• Specific Date Range (2017 - 2019) • Dissertations & Theses Global • Master’s and doctoral dissertation • All languages • English only search terms • Institutional subscription o Indexing 1743-present; Full text 1997-present o PQDT Global includes theses from Great Britain and Ireland	**76** [December 1, 2019]

Scopus: Search Strategy #3

Note: Advanced search selected to search within: title, abstract, and keywords.

Table S3. Metadata from Scopus Search Strategy #3

Search string	Restrictions	Returns [Date]
TITLE-ABS-KEY (Fish*) AND TITLE-ABS-KEY(“Fresh water” OR Freshwater OR Stream OR Water OR River OR Fluvial OR Estuar* OR Reservoir OR Impoundment OR "Hydro electric*" OR Hydroelectric* OR "Hydro dam*" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam) AND TITLE-ABS-KEY(Flow* OR Discharg*) AND TITLE-ABS-KEY(Productivity OR Biomass OR Abundance OR Densit* OR Yield OR “Ecological response” OR “Ecosystem response” OR “Biotic response”) AND NOT TITLE-ABS-KEY(mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon)	• 2017-2020 • Advanced search • All subject areas • All languages • All documents types • English only search terms	**774** [December 1, 2019]

Federal Science Library: Search Strategy #4
Available online: https://fsl-bsf.summon.serialssolutions.com/en/advanced#!/advanced?l=en; Federal Science Library (Formerly WAVES).

Note: Advanced search using Subject field searched words in the title, subject, series and abstract areas of the Federal Science Library record. Searching is a little more limited than previous databases.

Table S4. Metadata from Federal Science Library Search Strategy #4

Search string	Restrictions	Returns [Date]
(subjectTerms:(Fish*)) AND (subjectTerms:(Flow* OR Discharg*)) AND (subjectTerms:(Productivity OR Biomass OR Abundance OR Densit* OR Yield OR "Ecological response" OR "Ecosystem response" OR "Biotic response") AND (subjectTerms:(("Fresh water" OR Freshwater OR Stream OR Water OR River OR Fluvial OR Estuar* OR Reservoir OR Impoundment OR "Hydro electric" OR Hydroelectric* OR "Hydro dam" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam)) NOT (subjectTerms:(mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon))	• 2017 - 2019 • Advanced search • No language specified • Any content types • Any disciplines • No limits, no exclusions • Not expanded to outside your library's collection • Sort by Relevance • Open access – no institutional	**82** [December 1, 2019]

Science.gov: Search Strategy #5
Available online: https://www.science.gov/scigov/desktop/en/ostiblue/search.html

Note: Advanced search searches Full Record, Title, Author, and Date Range. You cannot specify abstract, so full record was searched. Search is more limited than previous databases.

Table S5. Metadata from Science.gov Search Strategy #5

Search string	Restrictions	Returns [Date]
Full Record: (Fish*) AND ("Fresh water" OR Freshwater OR Stream OR Water OR River OR Fluvial OR Estuar* OR Reservoir OR Impoundment OR "Hydro electric" OR Hydroelectric* OR "Hydro dam" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam) AND (Flow* OR Discharg*) AND (Productivity OR Biomass OR Abundance OR Densit* OR Yield OR "Ecological response" OR "Ecosystem response" OR "Biotic response") NOT (mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon)	• 2017-2019 • Advanced search • Sort by Relevance • All categories (except HSDB Hazardous Substances Databank) • Full record search terms • English only search terms • Accepted the additional records • Text records included only (no multimedia found) • Public access – peer-reviewed articles federally funded • Open access – no institutional subscription needed	**512** (+204 public access) [December 1, 2019]

AGRICOLA: Search Strategy #6
Available online: https://agricola.nal.usda.gov/vwebv/searchAdvanced

Note: Article Citation Database. Keywords Anywhere results come from anywhere in the bibliographical record: publisher, contents, note, meeting name, subject heading, sponsoring organization etc. Searching is more limited than other databases.

Table S6. Metadata from AGRICOLA Search Strategy #6

Search String	Restrictions	Returns [Date]
Fish? AND (Flow? OR Discharg?) AND (Productivity OR Biomass OR Abundance? OR Densit? OR Yield?)	• 2017-2019 • Searched NAL Article Citation Database • Advanced search • Used "all of these" • Searched within Keyword Anywhere • All Locations • All places • All types • All Formats • All languages • All media	**0** [October 11, 2019]

Search Engine
Internet searches were conducted in August 2019 using the search engine Google Scholar (first 500 hits sorted by relevance). Potentially useful documents that had not already been found in publication databases or in the systematic map were recorded and screened for inclusion in the review (Table S7).

Google Scholar: Search Strategy #7
Available online: https://scholar.google.com

Note: Keywords Anywhere results come from anywhere in the article. Searching is more limited than other databases. Number of actual returns exceeds reported, but limited to 500 most relevant.

Table S7. Metadata from Google Scholar Search Strategy #7 without using "Not" statements*

Search String	Restrictions	Returns [Date]
Find articles: • with **all** of the words Fish Flow • with **at least one** of the words: Productivity Biomass Abundance Density Yield Final search string appears as: Fish Flow Productivity OR Biomass OR abundance OR Density Or Yield	• 2017-2019 • Advanced search • Searched "where my words occur" – anywhere in the article • Search articles (excluding patents, case law, and citations) • Search for pages written in any language • Sort by relevance	**~82800** **Selected first 500** [January 8, 2020]

*The results of this search were included in the final count of Google Scholar articles

Other literature searches

A few specialist websites and databases of grey literature were suggested through our calls for literature that had previously not be identified or searched during the systematic map. These sites were searched January and March 2020.

ARLIS - Susitna Doc Finder: Search Strategy #8

Available online: https://www.arlis.org/susitnadocfinder/Search/Advanced; ARLIS – Alaska Resources Library and Information Services, Susitna Doc Finder – Online SuHydro and SuWa Documents

Note: Advanced Search: All Fields searches title, author, subject, publisher, year of publication, report number, full text and contents. Search uses search operators ALL, ANY or NO.

Table S8. Metadata from ARLIS Search Strategy #8

Search string	Restrictions	Returns [Date]
(All Fields:Fish*) AND (All Fields:"Fresh water" OR Freshwater OR Stream$ OR Water$ OR River$ OR Fluvial OR Estuar* OR Reservoir$ OR Impoundment$ OR "Hydro electric*" OR Hydroelectric* OR "Hydro dam*" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam$) AND (All Fields:Flow* OR Discharg*) AND (All Fields:Productivity OR Biomass OR Abundance$ OR Densit* OR Yield$ OR "Ecological response" OR "Ecosystem response" OR "Biotic response") NOT ((All Fields:mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon))	• 1904 onwards • Advanced search • Sort by Relevance • All authors, publishers, regions, topics, report types, report numbers • Full record search terms • Used Search Fields and Search Groups	11161 Selected first 250 [Jan 25, 2020]

Federal Energy Regulatory Commission (FERC) eLibrary: Search Strategy #9

Available online: https://elibrary-backup.ferc.gov/idmws/search/fercgensearch.asp; FERC Online eLibrary (formerly FERRIS) – Federal Energy Regulatory Commission eLibrary

Note: Advanced Search: Full Text searches rely upon the content of the FERC PDF, which may be incomplete. Accordingly, not all documents may be returned.

Table S9. Metadata from FERC Search Strategy #9

Search string	Restrictions	Returns [Date]
(Fish) AND ("Fresh water" OR Freshwater OR Stream OR Water OR River OR Fluvial OR Estuary OR Reservoir OR Impoundment OR "Hydro electric" OR Hydroelectric OR "Hydro dam" OR Hydrodam OR "Hydro power" OR Hydropower OR "Hydro" OR Dam) AND (Flow OR Discharg) AND (Productivity OR Biomass OR Abundance$ OR Density OR Yield OR "Ecological response" OR "Ecosystem response" OR "Biotic response") NOT (mining OR "mine site" OR aquaculture OR "wastewater treatment" OR carbon)	• Advanced search • Select Document Date: 01/01/1904 – 12/31/2019 • Search Library: Hydro • Full Text and Description search terms • Document Type for class and type: All • Search only Public availability • Score Desc	**11161** **Selected only first 250** [March 16, 2020]

Other literature searches

Reference sections of accepted articles and relevant reviews were hand searched to evaluate relevant titles that were not found using the search strategy. Stakeholders were consulted for insight and advice for new sources of information. We also issued a call for evidence to target sources of grey literature through relevant mailing lists (CCFFR, CHA, CEA, OWA, Western Division American Fisheries Society, Mactaquac Aquatic Ecosystem Study, IFC) and through social media (e.g., Twitter) in December 2019. The call for evidence was also distributed by the Advisory Team to relevant networks and colleagues.

Original Search String – Systematic Map

Table S10. Original search string used to in the systematic map (Rytwinski et al. 2017).

Component	Search string
Population terms	[Fish* AND ("Fresh water" OR Freshwater OR Stream$ OR Water$ OR River$ OR Fluvial OR Lake$ OR Pond$ OR Wetland$ OR Estuar* OR Reservoir$ OR Canal$ OR Impoundment$ OR "Hydro electric*" OR Hydroelectric* OR "Hydro dam*" OR Hydrodam* OR "Hydro power" OR Hydropower OR "Hydro" OR Dam$ OR Withdraw* OR Diversion$ OR "Climate change")]
	AND
Intervention/exposure terms	(Flow* OR Discharg*)
	AND
Outcome terms	(Productivity OR Growth OR Performance OR Surviv* OR Success OR Migrat* OR Passag* OR Reproduc* OR Biomass OR Stress* OR Disease$ OR Mortalit* OR Abundance$ OR Densit* OR Recruit* OR Yield$ OR "Ecological response" OR "Ecosystem response" OR "Biotic response")

Appendix 4. Excluded articles

Description: List of articles excluded on the basis of full-text assessment or data extraction and reasons for exclusion. Separate lists of articles excluded on the basis of full-text assessment, assessment during data-extraction, articles that were unobtainable and relevant reviews. See OSF link in section "Availability of data and materials".

Appendix 5. Data-extraction sheet

Description: Data-extraction sheet. Contains the coding (extracted data) for all articles/studies included in the narrative synthesis. Includes a description of the coding form, the actual coding of all articles/studies, and a list of supplementary articles. See OSF link in section "Availability of data and materials".

Appendix 6. Outcome metric decision tree

Description: Outcome metric decision tree. Includes decision tree for (a) abundance (Fig. S1) and (b) biomass. (Fig. S2). Used when selecting outcome metrics for quantitative synthesis.

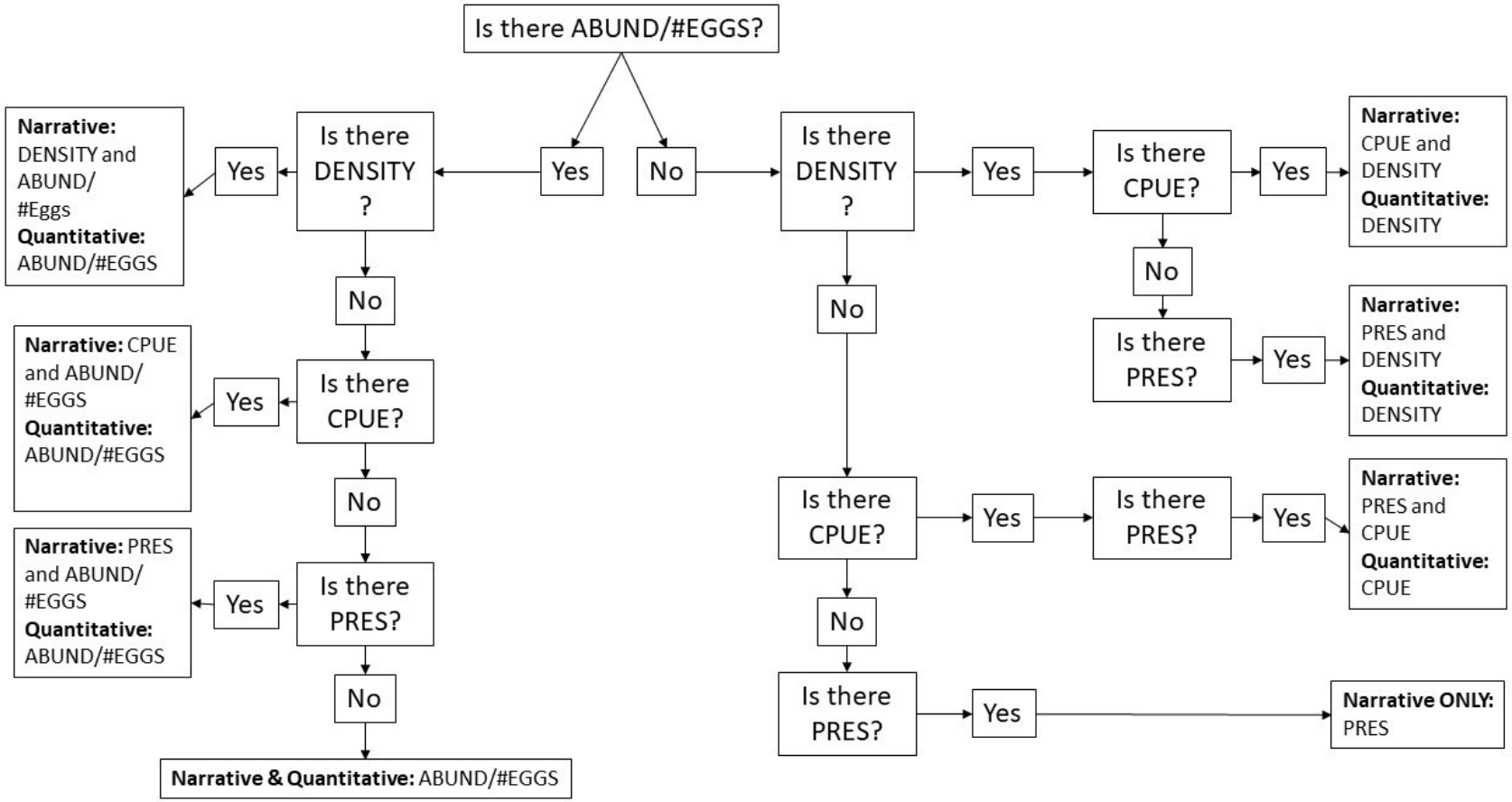

Fig. S1. Decision tree for abundance metrics to assist in determining whether to retain metrics for narrative and/or quantitative synthesis. ABUND: abundance; CPUE: catch per unit effort; DENSITY: density; #EGGS: number of eggs; PRES: presence/absence.

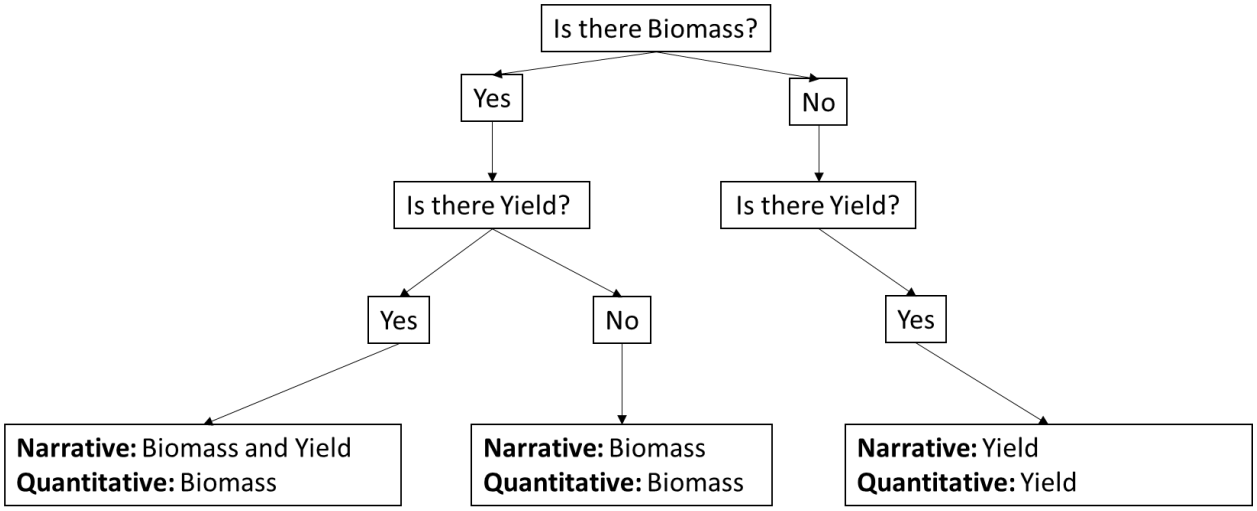

Fig. S2. Decision tree for biomass metrics to assist in determining whether to retain metrics for narrative and/or quantitative synthesis.

This space intentionally left blank.

Appendix 7. Data preparation for quantitative synthesis and additional calculations for meta-analysis

Description: Here, we provide a description of data preparation for quantitative synthesis in relation to reducing multiple effect size estimates from the same study, and our handling of pseudoreplication.

Combining data across multiple comparisons within a study

To reduce multiple effect size estimates from the same study, and avoid giving studies with multiple estimates more weight in analyses, datasets were aggregated in five instances when studies sharing all other meta-data, reported: (i) responses from multiple life stages separately within the same outcome (e.g., the abundance of eggs for species X, and the abundance of age-0 for species X, separately) (seven studies); (ii) years and/or seasons post-treatment within a given outcome category (i.e., if for a given outcome category, multiple seasons were monitored and reported separately for a *CI* study design (one study; not included for meta-analysis due to lack of replication) or within-in year variation post-treatment for a *BA* design) (four studies); (iii) different sampling methods (no studies), (iv) sites downstream of a hydro dam within a single river sampled using a *BA* design (13 studies), and (v) for one study in which data for both resident and non-resident individuals of the same species were reported and aggregated. Responses could have been summed if sampling was conducted in the same time period; however, in all instances where this occurred, aggregation would have led to an inflated sample size and required an unacceptable level of data manipulation.

When aggregation was necessary, in each of the cases discussed above, we computed an average effect size. To do so, we first computed the arithmetic mean effect size as the mean of the effect sizes from different comparisons within a study according to equation 24.1 in Borenstein et al. (2009c). The variance of this average effect size was then computed following

methods of Borenstein et al. (2009c). Because the correlations between comparisons were unknown, we assumed a correlation coefficient of r=1. This assumption may lead to overestimation of variance and result in an increased likelihood of a Type II error (i.e., finding that the effect size is not significantly different from zero). Given our small database of effect sizes for quantitative analysis in *CI* and within-year *BA* studies, and our relatively small sample size for interannual *BA* studies, we were limited in our ability to use other approaches such as multivariate or robust variance estimation (Hedges et al. 2010; Tanner-Smith and Tipton 2014).

Adjustment accounting for pseudoreplication

Replication within a *control/impact* study (i.e., group sample size) was considered at two levels: (i) independent intervention areas (i.e., separate waterbodies or separate sections of a waterbody receiving treatment – true replicates), and (ii) 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 (i) 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 sub-sample site 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.

To avoid giving pseudoreplicated data too much weight in analyses (i.e. outcome means and variances were not from independent replicates but subsampled sites), we calculated the variance of effect sizes using a modified equation and a conservative sample number (Bernes et al. 2018; Eales et al. 2018).

1. Standard errors of the mean outcomes for each group were converted to standard

deviations using the total numbers of subsamples as sample sizes (or reported standard deviations based on the total number of subsamples that were left unchanged).

2. Effect sizes (Hedges' *g* statistic) that were also calculated using the total number of subsamples (Borenstein et al. 2009b):

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

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:

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

where S = standard deviation, and n_{G2} and n_{G2} were the number of subsamples of group 1 and group 2 (e.g., number of subplots or nests).

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

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

here again, n_{G1} and n_{G2} were the number of subsamples of group 1 and group 2.

3. In the following equation, variances for *d* (i.e., V_d) were calculated using both the conservative sample number (solid-lined box; nt_{G1} and nt_{G2}) and the total number of subsamples as sample sizes (dashed-lined box; n_{G1} and n_{G2}):

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

This conservative sample number was based on using the number of true replicates instead of pseudoreplicates. For this review, in all cases, the conservative sample number was one replicate waterbody (i.e., $nt_{G1}=1$ and $nt_{G2}=1$) where plot or nest subsamples were taken from.

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

$$Hedge's\ g = J \times d$$

$$V_g = J^2 \times V_d$$

Therefore, the total number of subsamples was used to calculate an effect size estimate and the conservative sample number was used to calculate the uncertainty of this estimate (i.e., effect size weighting). This method provides a conservative estimate of variability. We used sensitivity analyses to determine the influence of including pseudoreplicated studies on the overall mean effect size.

Missing Variance

When standard deviations were missing and it was not possible to calculate from reported data (i.e., by calculating from reported raw data, converting from standard error or 95% confidence interval), mean value imputation was used to calculate the missing variables:

$$SD = \bar{X}_J \left(\frac{\sum_{G2}^{k} SD_{G2}}{\sum_{G2}^{K} \overline{X_{G2}}} \right)$$

where $\bar{X}_j$ is the observed mean of the dataset with missing information, and k is the number of j studies with complete information. All datasets with standard deviations were used in

imputation, unless effect sizes were calculated through aggregation. Imputation methods have been found to perform as well, or better, than other ways of dealing with missing data values as long as the datasets with missing variables did not exceed 60% of the total (Kambach et al. 2020). In our case, two *CI* studies (eight cases, of which six cases were aggregated resulting in three datasets) had missing standard deviations that could not be calculated from reported data. This is equal to 3% of all *CI* study datasets, allowing us to confidently use imputation to fill these data gaps. To determine the impact of imputation, we compared the summary effect sizes with and without datasets with imputation using sensitivity analysis.

Appendix 8. Study validity assessment

Description: This file contains a description of the study validity assessment tool and results of assessments for each article/study included in the narrative synthesis. See OSF link in section "Availability of data and materials".

Appendix 9. Fish species list

Description: Includes all family, genera and species included in the narrative synthesis.

Table S1. Fish species list with all family, genera and species included in the narrative synthesis.

Family	Genus	Species	Common name
Acestrorhynchidae	Acestrorhynchus	*Acestrorhynchus pantaneiro*	[unk]
Acipenseridae	Acipenser	*Acipenser fulvescens*	Lake sturgeon
		Acipenser sinensis	Chinese sturgeon
Amiidae	Amia	*Amia calva*	Bowfin
Anguillidae	Anguilla	*Anguilla anguilla*	European eel
		Anguilla australis	Short-finned eel
		Anguilla reinhardtii	Speckled longfin eel
		Anguilla rostrata	American eel
Anostomidae	Leporinus	*Leporinus amae*	[unk]
	Schizodon	*Schizodon nasutus*	[unk]
Aphredoderidae	Aphredoderus	*Aphredoderus sayanus*	Pirate perch
Atherinopsidae	Labidesthes	*Labidesthes sicculus*	Brook silverside
Belonidae	Strongylura	*Strongylura marina*	Atlantic needlefish
Catostomidae	Carpiodes	*Carpiodes carpio*	River carpsucker
		Carpiodes cyprinus	Quillback
		Carpiodes velifer	Highfin carpsucker
	Catostomus	*Catostomus catostomus*	Longnose sucker
		Catostomus columbianus	Bridgelip sucker
		Catostomus commersonii	White sucker
		Catostomus discobolus	Bluehead sucker
		Catostomus latipinnis	Flannelmouth sucker
		Catostomus macrocheilus	Largescale sucker
		Catostomus occidentalis	Sacramento sucker
		Catostomus platyrhynchus	Mountain sucker
	Cycleptus	*Cycleptus elongatus*	Blue sucker
	Erimyzon	*Erimyzon oblongus*	Eastern creek chubsucker
	Hypentelium	*Hypentelium etowanum*	Alabama hog sucker
		Hypentelium nigricans	Northern hog sucker
	Ictiobus	*Ictiobus bubalus*	Smallmouth buffalo
		Ictiobus cyprinellus	Bigmouth buffalo
	Minytrema	*Minytrema melanops*	Spotted sucker
	Moxostoma	*Moxostoma anisurum*	Silver redhorse
		Moxostoma breviceps	Smallmouth redhorse
		Moxostoma carinatum	River redhorse
		Moxostoma collapsum	Notchlip redhorse
		Moxostoma duquesnei	Black redhorse
		Moxostoma erythrurum	Golden redhorse
		Moxostoma macrolepidotum	Shorthead redhorse
		Moxostoma pappillosum	V-lip redhorse
		Moxostoma pisolabrum	Pealip redhorse
		Moxostoma poecilurum	Blacktail redhorse
	Xyrauchen	*Xyrauchen texanus*	Razorback sucker
Centrarchidae	Ambloplites	*Ambloplites ariommus*	Shadow bass
		Ambloplites rupestris	Rock bass
	Centrarchus	*Centrarchus macropterus*	Flier
	Lepomis	*Lepomis auritus*	Redbreast sunfish
		Lepomis cyanellus	Green sunfish
		Lepomis gibbosus	Pumpkinseed

Family	Genus	Species	Common name
		Lepomis gulosus	Warmouth
		Lepomis humilis	Orangespotted sunfish
Centrarchidae	Lepomis	*Lepomis macrochirus*	Bluegill
		Lepomis megalotis	Longear sunfish
		Lepomis microlophus	Redear sunfish
		Lepomis miniatus	Redspotted sunfish
		Lepomis punctatus	Spotted sunfish
	Micropterus	*Micropterus coosae*	Redeye bass
		Micropterus dolomieu	Smallmouth bass
		Micropterus punctulatus	Spotted bass
		Micropterus salmoides	Largemouth black bass
	Pomoxis	*Pomoxis annularis*	White crappie
		Pomoxis nigromaculatus	Black crappie
Characidae	Astyanax	*Astyanax fasciatus*	Banded astyanax
		Astyanax gr. Scabripinnis	[unk]
		Astyanax jacuhiensis	[unk]
	Bryconamericus	*Bryconamericus iheringii*	[unk]
		Bryconamericus stramineus	[unk]
	Oligosarcus	*Oligosarcus jenynsii*	[unk]
Clupeidae	Alosa	*Alosa aestivalis*	Blueback shad
		Alosa alabamae	Alabama shad
		Alosa chrysochloris	Skipjack shad
		Alosa pseudoharengus	Alewife
		Alosa sapidissima	American shad
	Dorosoma	*Dorosoma cepedianum*	American gizzard shad
		Dorosoma petenense	Threadfin shad
	Potamalosa	*Potamalosa richmondia*	Freshwater herring
Cobitidae	Cobitis	*Cobitis maroccana*	[unk]
		Cobitis taenia	Spined loach
	Misgurnus	*Misgurnus fossilis*	Weatherfish
	Sinibotia	*Sinibotia superciliaris*	[unk]
Cottidae	Cottus	*Cottus asper*	Prickly sculpin
		Cottus bairdii	Mottled sculpin
		Cottus caeruleomentum	Blue ridge sculpin
		Cottus carolinae	Banded sculpin
		Cottus cognatus	Slimy sculpin
		Cottus confusus	Shorthead Sculpin
		Cottus gobio	Bullhead
		Cottus hubbsi	Columbia Sculpin
		Cottus hypselurus	Ozark sculpin
		Cottus klamathensis	Marbled sculpin
		Cottus pitensis	Pit sculpin
		Cottus ricei	Spoonhead sculpin
		Cottus tallapoosae	[unk]
Curimatidae	Steindachnerina	*Steindachnerina sp.*	[unk]
Cyprinidae	Abramis	*Abramis brama*	Freshwater bream
		Abramis sapa	White-eye bream
	Achondrostoma	*Achondrostoma arcasii*	[unk]
	Alburnoides	*Alburnoides bipunctatus*	Schneider
	Alburnus	*Alburnus alburnus*	Bleak
	Aspius	*Aspius sp.*	[unk]
	Ballerus	*Ballerus ballerus*	Zope

Family	Genus	Species	Common name
	Barbus	*Barbus barbus*	Barbel
	Blicca	*Blicca bjoerkna*	White bream
Cyprinidae	Campostoma	*Campostoma anomalum*	Central stoneroller
		Campostoma oligolepis	Largescale stoneroller
	Carassius	*Carassius auratus*	Goldfish
	Chondrostoma	*Chondrostoma nasus*	Common nase
	Chrosomus	*Chrosomus erythrogaster*	Southern redbelly dace
	Clinostomus	*Clinostomus funduloides*	Rosyside dace
	Coreius	*Coreius heterodon*	[unk]
	Couesius	*Couesius plumbeus*	Lake chub
	Ctenopharyngodon	*Ctenopharyngodon idella*	Grass carp
	Cyprinella	*Cyprinella analostana*	Satinfin shiner
		Cyprinella callistia	Alabama shiner
		Cyprinella gibbsi	Tallapoosa shiner
		Cyprinella lutrensis	Red shiner
		Cyprinella spiloptera	Spotfin shiner
		Cyprinella venusta	Blacktail shiner
		Cyprinella whipplei	Steelcolor shiner
	Cyprinus	*Cyprinus carpio*	Common carp
	Erimystax	*Erimystax dissimilis*	Streamline chub
		Erimystax x-punctatus	Gravel chub
	Exoglossum	*Exoglossum maxillingua*	Cutlips minnow
	Gila	*Gila atraria*	Utah chub
		Gila cypha	Humpback chub
		Gila robusta	Roundtail chub
	Gobio	*Gobio gobio*	Gudgeon
	Hesperoleucus	*Hesperoleucus symmetricus*	California roach
	Hybognathus	*Hybognathus hayi*	Cypress minnow
		Hybognathus regius	Eastern silvery minnow
	Hybopsis	*Hybopsis winchelli*	Clear chub
	Hypophthalmichthys	*Hypophthalmichthys molitrix*	Silver carp
		Hypophthalmichthys nobilis	Bighead carp
	Leuciscus	*Leuciscus aspius*	Asp
		Leuciscus idus	Ide
		Leuciscus leuciscus	Common dace
	Luxilus	*Luxilus chrysocephalus*	Striped shiner
		Luxilus cornutus	Common shiner
		Luxilus zonatus	Bleeding shiner
	Lythrurus	*Lythrurus bellus*	Pretty shiner
		Lythrurus fumeus	Ribbon shiner
		Lythrurus umbratilis	Redfin shiner
	Macrhybopsis	*Macrhybopsis aestivalis*	Speckled chub
		Macrhybopsis storeriana	Silver chub
	Mylocheilus	*Mylocheilus caurinus*	Peamouth
	Mylopharodon	*Mylopharodon conocephalus*	Hardhead
	Mylopharyngodon	*Mylopharyngodon piceus*	Black carp
	Nocomis	*Nocomis biguttatus*	Hornyhead chub
		Nocomis leptocephalus	Bluehead chub
		Nocomis micropogon	River chub
	Notemigonus	*Notemigonus crysoleucas*	Golden shiner
	Notropis	*Notropis ammophilus*	Orangefin shiner
		Notropis amoenus	Comely shiner

Family	Genus	Species	Common name
		Notropis atherinoides	Emerald shiner
		Notropis baileyi	Rough shiner
Cyprinidae	Notropis	*Notropis boops*	Bigeye shiner
		Notropis buccatus	Silverjaw minnow
		Notropis buchanani	Ghost shiner
		Notropis candidus	Silverside shiner
		Notropis edwardraneyi	Fluvial shiner
		Notropis greenei	Wedgespot shiner
		Notropis hudsonius	Spottail shiner
		Notropis nubilus	Ozark minnow
		Notropis percobromus	Carmine shiner
		Notropis procne	Swallowtail shiner
		Notropis rubellus	Rosyface shiner
		Notropis stilbius	Silverstripe shiner
		Notropis stramineus	Sand shiner
		Notropis texanus	Weed shiner
		Notropis uranoscopus	Skygazer shiner
		Notropis volucellus	Mimic shiner
	Opsopoeodus	*Opsopoeodus emiliae*	Pugnose minnow
	Parachondrostoma	*Parachondrostoma toxostoma*	[unk]
	Phenacobius	*Phenacobius catostomus*	Riffle minnow
	Phoxinus	*Phoxinus phoxinus*	Eurasian minnow
	Pimephales	*Pimephales notatus*	Bluntnose minnow
		Pimephales promelas	Fathead minnow
		Pimephales vigilax	Bullhead minnow
	Platygobio	*Platygobio gracilis*	Flathead chub
	Ptychocheilus	*Ptychocheilus grandis*	Sacramento pikeminnow
		Ptychocheilus lucius	Colorado pikeminnow
		Ptychocheilus oregonensis	Northern pikeminnow
	Rhinichthys	*Rhinichthys atratulus*	Blacknose dace
		Rhinichthys cataractae	Longnose dace
		Rhinichthys osculus	Speckled dace
		Rhinichthys umatilla	Umatilla Dace
	Rhodeus	*Rhodeus amarus*	European bitterling
		Rhodeus sericeus	Bitterling
	Richardsonius	*Richardsonius balteatus*	Redside shiner
	Romanogobio	*Romanogobio albipinnatus*	White-finned gudgeon
	Rutilus	*Rutilus rutilus*	Roach
	Scardinius	*Scardinius erythrophthalmus*	Rudd
	Schizothorax	*Schizothorax plagiostomus*	[unk]
	Semotilus	*Semotilus atromaculatus*	Creek chub
		Semotilus corporalis	Fallfish
	Squalius	*Squalius cephalus*	Chub
	Telestes	*Telestes souffia*	Vairone
	Tinca	*Tinca tinca*	Tench
	Vimba	*Vimba vimba*	Vimba bream
Diplomystidae	Diplomystes	*Diplomystes nahuelbutaensis*	[unk]
Elassomatidae	Elassoma	*Elassoma zonatum*	Banded pygmy sunfish
Eleotridae	Gobiomorphus	*Gobiomorphus australis*	Striped gudgeon
		Gobiomorphus coxii	Cox's gudgeon
	Hypseleotris	*Hypseleotris compressa*	Empire gudgeon
		Hypseleotris kluzingeri	Western carp gudgeon

Family	Genus	Species	Common name
	Philypnodon	*Philypnodon grandiceps*	Flat-headed gudgeon
Embiotocidae	Hysterocarpus	*Hysterocarpus traskii*	Russian river tule perch
Erythrinidae	Hoplias	*Hoplias sp.*	[unk]
Esocidae	Esox	*Esox americanus*	Redfin pickerel
		Esox lucius	Northern pike
		Esox niger	Chain pickerel
Fundulidae	Fundulus	*Fundulus catenatus*	Northern studfish
		Fundulus diaphanus	Banded killifish
		Fundulus olivaceus	Blackspotted topminnow
		Fundulus zebrinus	Plains killifish
Galaxiidae	Galaxias	*Galaxias brevipinnis*	Koaro
		Galaxias maculatus	Inanga
		Galaxias olidus	Mountain galaxias
Gasterosteidae	Culaea	*Culaea inconstans*	Brook stickleback
	Gasterosteus	*Gasterosteus aculeatus*	Three-spined stickleback
Gobiidae	Ponticola	*Ponticola kessleri*	Bighead goby
	Proterorhinus	*Proterorhinus marmoratus*	Tubenose goby
Gymnotidae	Gymnotus	*Gymnotus carapo*	Banded knifefish
Heptapteridae	Pimelodella	*Pimelodella sp.*	[unk]
	Rhamdia	*Rhamdia quelen*	South American catfish
Hiodontidae	Hiodon	*Hiodon alosoides*	Goldeye
		Hiodon tergisus	Mooneye
Ictaluridae	Ameiurus	*Ameiurus catus*	White catfish
		Ameiurus melas	Black bullhead
		Ameiurus natalis	Yellow bullhead
		Ameiurus nebulosus	Brown bullhead
		Ameiurus platycephalus	Flat bullhead
	Ictalurus	*Ictalurus furcatus*	Blue catfish
		Ictalurus punctatus	Channel catfish
	Noturus	*Noturus eleutherus*	Mountain madtom
		Noturus exilis	Slender madtom
		Noturus flavus	Stonecat
		Noturus funebris	Black madtom
		Noturus gyrinus	Tadpole madtom
		Noturus insignis	Margined madtom
		Noturus lachneri	Ouachita madtom
		Noturus leptacanthus	Speckled madtom
		Noturus nocturnus	Freckled madtom
	Pylodictis	*Pylodictis olivaris*	Flathead catfish
Lepisosteidae	Lepisosteus	*Lepisosteus oculatus*	Spotted gar
		Lepisosteus osseus	Longnose gar
		Lepisosteus platostomus	Shortnose gar
Loricariidae	Hypostomus	*Hypostomus sp.*	[unk]
Lotidae	Lota	*Lota lota*	Burbot
Mordaciidae	Mordacia	*Mordacia mordax*	Shorthead lamprey
Moronidae	Morone	*Morone americana*	White perch
Moronidae	Morone	*Morone chrysops*	White bass
		Morone chrysops x Morone saxitilis	[unk]
		Morone saxatilis	Striped bass
Mugilidae	Mugil	*Mugil cephalus*	Flathead grey mullet
	Trachystoma	*Trachystoma petardi*	Pinkeye mullet
Nemacheilidae	Barbatula	*Barbatula barbatula*	Stone loach

Family	Genus	Species	Common name
	Triplophysa	*Triplophysa marmorata*	[unk]
Osmeridae	Osmerus	*Osmerus mordax*	Rainbow smelt
Paralichthyidae	Paralichthys	*Paralichthys lethostigma*	Southern flounder
Parodontidae	Apareiodon	*Apareiodon affinis*	Darter characine
Percichthyidae	Macquaria	*Macquaria australasica*	Macquarie perch
		Macquaria colonorum	Estuary perch
		Macquaria novemaculeata	Australian bass
Percidae	Ammocrypta	*Ammocrypta beanii*	Naked sand darter
		Ammocrypta meridiana	Southern sand darter
	Crystallaria	*Crystallaria asprella*	Crystal darter
	Etheostoma	*Etheostoma blennioides*	Greenside darter
		Etheostoma caeruleum	Rainbow darter
		Etheostoma chuckwachatte	Lipstick darter
		Etheostoma collettei	Creole darter
		Etheostoma flabellare	Fantail darter
		Etheostoma gracile	Slough darter
		Etheostoma histrio	Harlequin darter
		Etheostoma nigrum	Johnny darter
		Etheostoma olmstedi	Tessellated darter
		Etheostoma proeliare	Cypress darter
		Etheostoma radiosum	Orangebelly darter
		Etheostoma rupestre	Rock darter
		Etheostoma spectabile	Orangethroat darter
		Etheostoma stigmaeum	Speckled darter
		Etheostoma tallapoosae	Tallapoosa darter
		Etheostoma tetrazonum	Missouri saddled darter
		Etheostoma variatum	Variegate darter
		Etheostoma whipplei	Redfin darter
		Etheostoma zonale	Banded darter
	Gymnocephalus	*Gymnocephalus baloni*	Danube ruffe
		Gymnocephalus cernua	Ruffe
	Perca	*Perca flavescens*	American yellow perch
		Perca fluviatilis	European perch
	Percina	*Percina bimaculata*	Chesapeake logperch
		Percina caprodes	Logperch
		Percina copelandi	Channel darter
		Percina cymatotaenia	Bluestripe darter
		Percina fulvitaenia	Ozark logperch
		Percina kathae	Mobile logperch
		Percina lenticula	Freckled darter
		Percina maculata	Blackside darter
		Percina nasuta	Longnose darter
		Percina nigrofasciata	Blackbanded darter
		Percina palmaris	Bronze darter
Percidae	Percina	*Percina peltata*	Shield darter
		Percina phoxocephala	Slenderhead darter
		Percina roanoka	Roanoke darter
		Percina sciera	Dusky darter
		Percina shumardi	River darter
		Percina smithvanizi	Muscadine Darter
		Percina uranidea	Stargazing darter
		Percina vigil	Saddleback darter

Family	Genus	Species	Common name
	Sander	*Sander canadensis*	Sauger
		Sander lucioperca	Pike-perch
Percidae	Sander	*Sander vitreus*	Walleye
Perciliidae	Percilia	*Percilia irwini*	[unk]
Percopsidae	Percopsis	*Percopsis omiscomaycus*	Trout-perch
Petromyzontidae	Entosphenus	*Entosphenus lethophagus*	Pit-Klamath brook lamprey
	Eudontomyzon	*Eudontomyzon mariae*	Ukrainian brook lamprey
	Ichthyomyzon	*Ichthyomyzon castaneus*	Chestnut lamprey
		Ichthyomyzon fossor	Northern brook lamprey
		Ichthyomyzon gagei	Southern brook lamprey
	Lampetra	*Lampetra planeri*	European brook lamprey
		Lampetra richardsoni	Western brook lamprey
	Petromyzon	*Petromyzon marinus*	Sea lamprey
Pimelodidae	Parapimelodus	*Parapimelodus valenciennis*	[unk]
	Pimelodus	*Pimelodus absconditus*	[unk]
		Pimelodus atrobrunneus	[unk]
		Pimelodus maculatus	[unk]
Plotosidae	Tandanus	*Tandanus tandanus*	Freshwater catfish
Poeciliidae	Gambusia	*Gambusia affinis*	Mosquitofish
		Gambusia holbrooki	Eastern mosquitofish
Polyodontidae	Polyodon	*Polyodon spathula*	Mississippi paddlefish
Pseudomugilidae	Pseudomugil	*Pseudomugil signifer*	Pacific blue eye
Retropinnidae	Prototroctes	*Prototroctes maraena*	Australian grayling
	Retropinna	*Retropinna semoni*	Australian smelt
Salmonidae	Coregonus	*Coregonus artedi*	Cisco
		Coregonus clupeaformis	Lake whitefish
	Hucho	*Hucho hucho*	Huchen
	Oncorhynchus	*Oncorhynchus clarkii*	Cutthroat trout
		Oncorhynchus clarkii clarkii	Coastal Cutthroat Trout
		Oncorhynchus gorbuscha	Pink salmon
		Oncorhynchus keta	Chum salmon
		Oncorhynchus kisutch	Coho salmon
		Oncorhynchus mykiss	Rainbow trout
		Oncorhynchus nerka	Sockeye salmon
		Oncorhynchus tshawytscha	Chinook salmon
	Prosopium	*Prosopium williamsoni*	Mountain whitefish
	Salmo	*Salmo salar*	Atlantic salmon
		Salmo trutta	Sea trout
	Salvelinus	*Salvelinus confluentus*	Bull trout
		Salvelinus fontinalis	Brook trout
		Salvelinus leucomaenis	Whitespotted char
		Salvelinus malma	Dolly varden
Salmonidae	Thymallus	*Thymallus thymallus*	Grayling
Sciaenidae	Aplodinotus	*Aplodinotus grunniens*	Freshwater drum
Siluridae	Silurus	*Silurus glanis*	Wels catfish
Sisoridae	Glyptothorax	*Glyptothorax pectinopterus*	River cat
Sparidae	Acanthopagrus	*Acanthopagrus australis*	Yellowfin bream
Sternopygidae	Eigenmannia	*Eigenmannia virescens*	Glass knifefish
Tetrarogidae	Notesthes	*Notesthes robusta*	Bullrout
Trichomycteridae	Trichomycterus	*Trichomycterus areolatus*	[unk]

Appendix 10. Quantitative synthesis database

Description: Contains the coding (extracted data) for all articles/studies/datasets included in the quantitative synthesis. Includes the actual coding of all articles/studies/datasets, the calculations of effect sizes, calculations for aggregating effect sizes and a list of articles/studies/datasets not considered during quantitative meta-analysis. See OSF link in section "Availability of data and materials".

Appendix 11. Global meta-analyses and publication bias

Description: Global meta-analyses, publication bias and moderator analysis. All forest (i.e, summary plot of all effect size estimates) and funnel (i.e., visual assessment of publication bias using a scatter plot of effect sizes versus a measure of precision) plots from global and sensitivity analyses. Correlation analyses of moderator (Pearson's χ^2) and additional meta-analysis calculations and results are also reported.

This space intentionally left blank

Global meta-analysis – *Control/Impact* Studies
Abundance

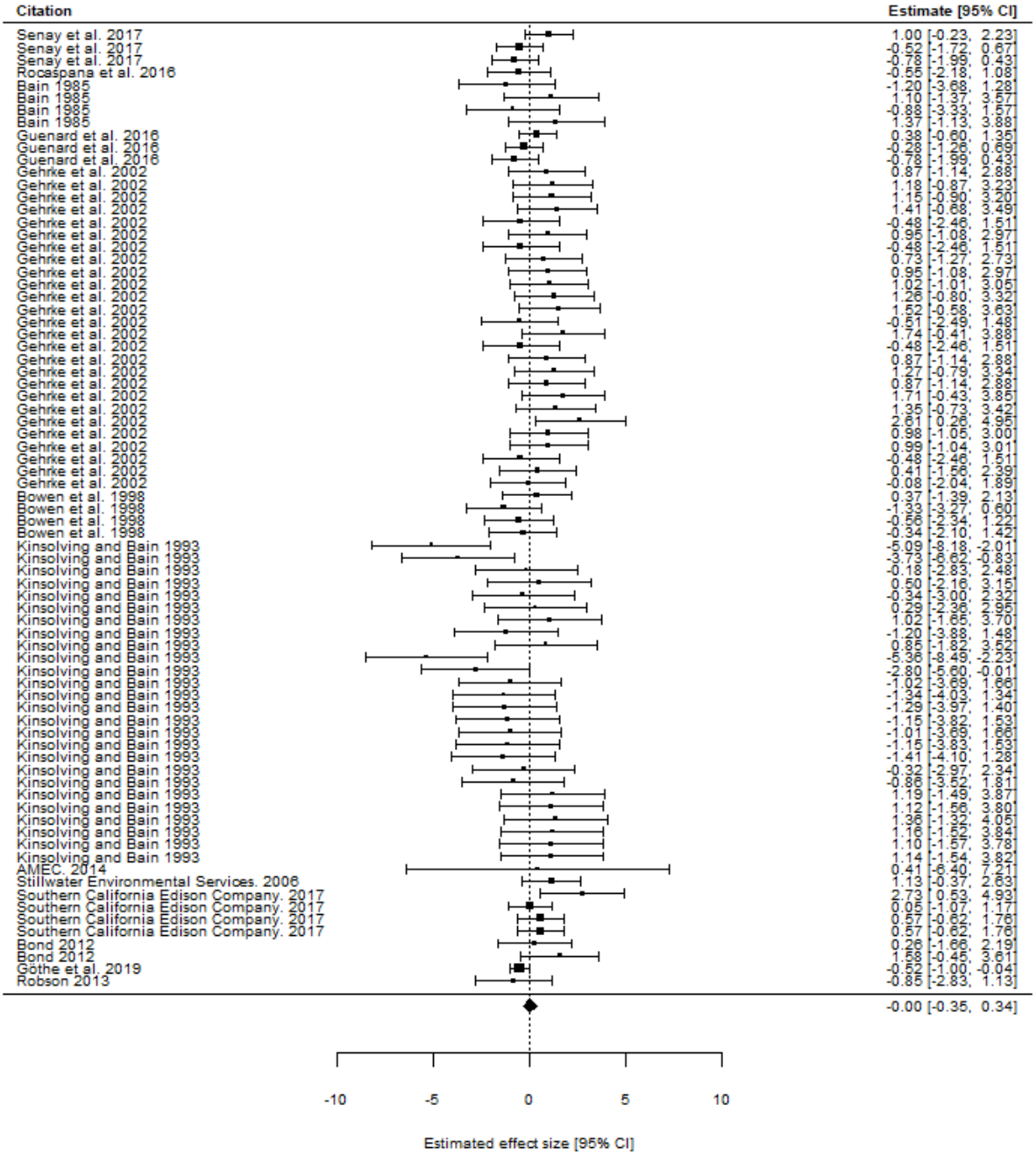

Fig. S1. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on fish abundance (*k*=77). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). Diamond: overall mean effect size of random-effects model.

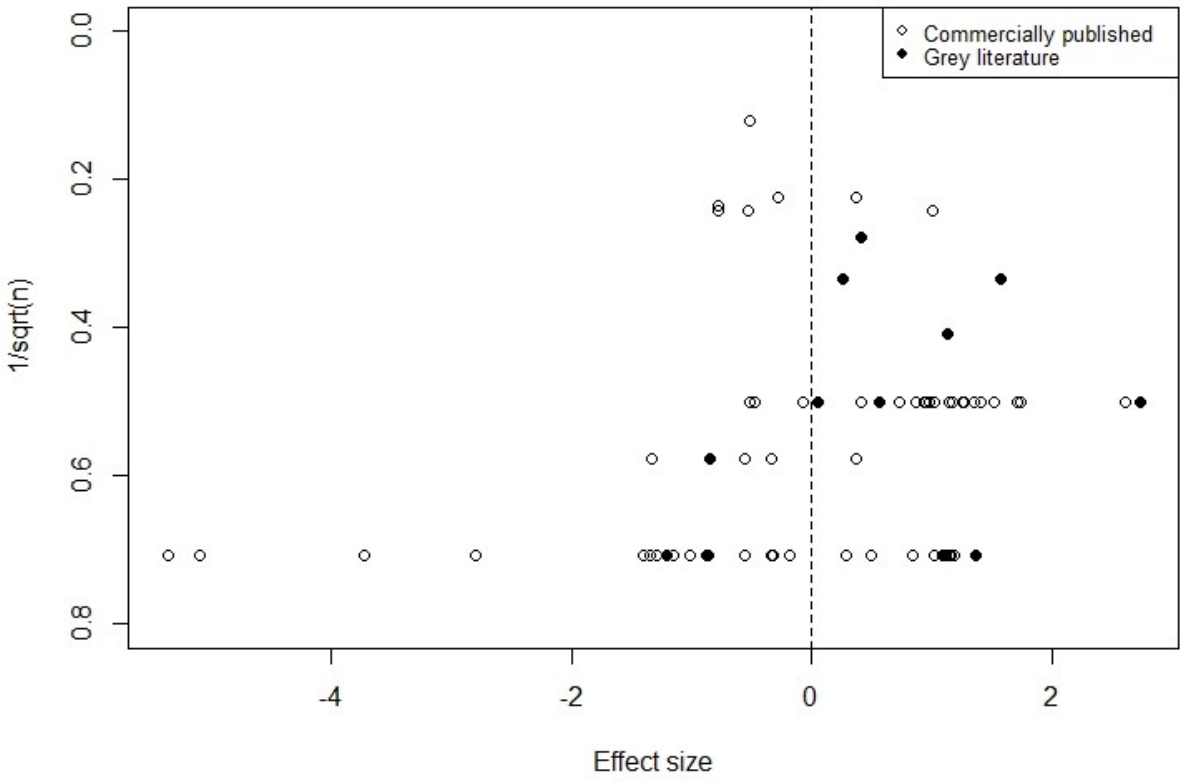

Fig. S2 Funnel plot showing *Control/Impact* studies of abundance (k = 77). Open circles indicate datasets from commercially published articles and filled circles indicate datasets from grey literature. Summary effect is indicated by the dashed line.

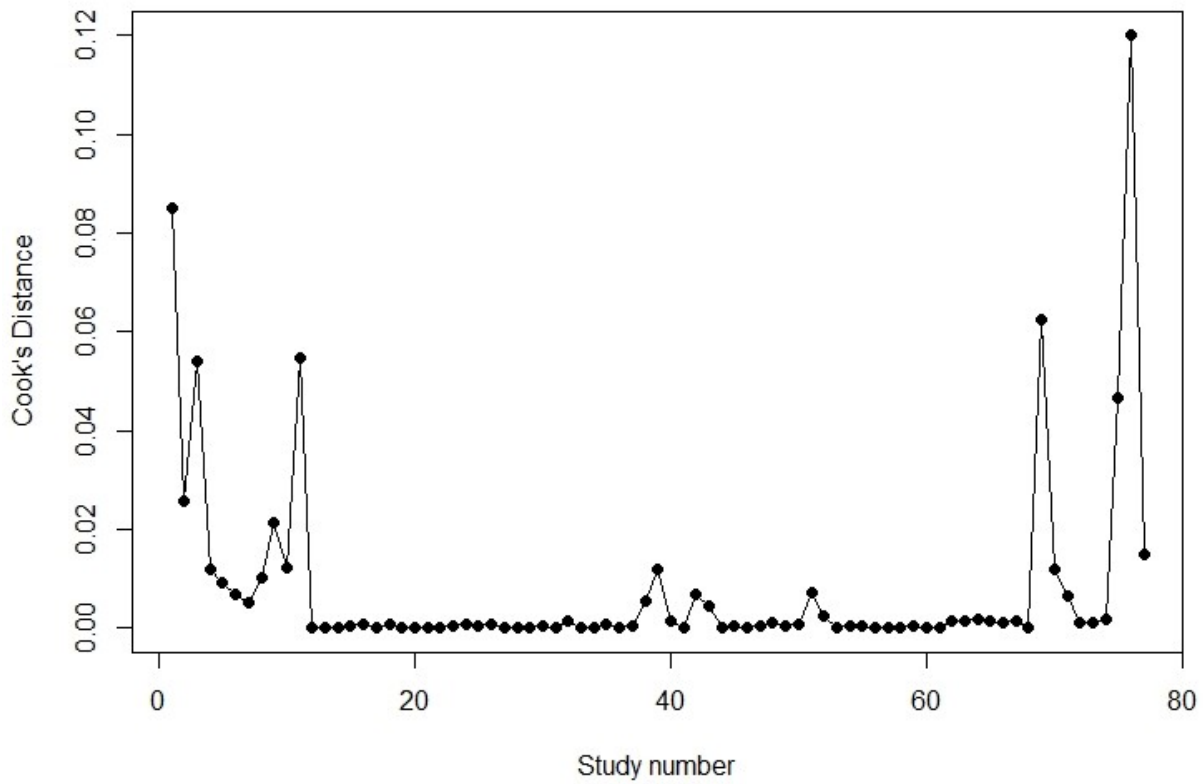

Fig. S3. Cook's distance plot indicating influence of effect size. Note the outlier of concern (Cook's distance ≈ 0.12).

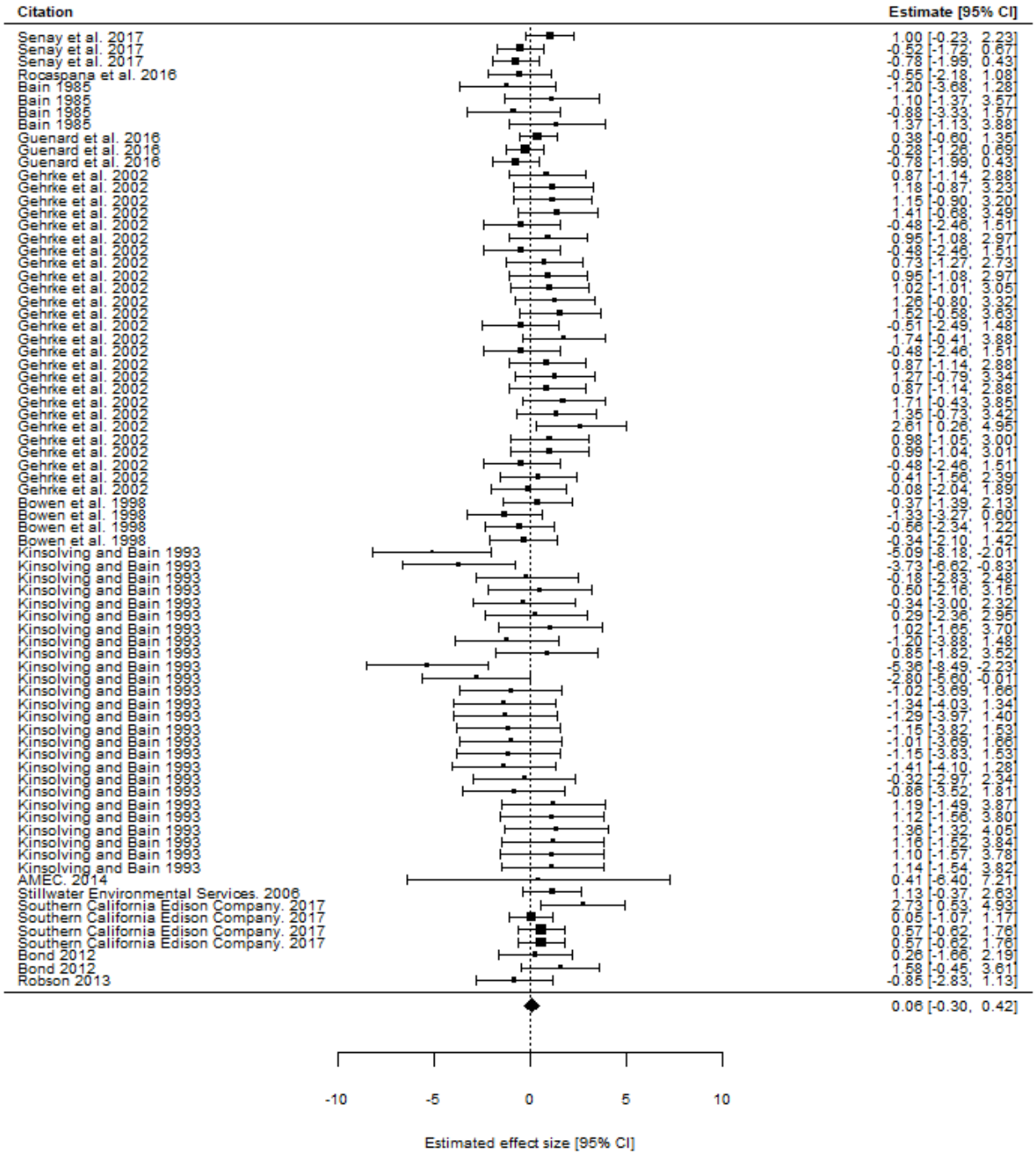

Fig. S4. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on fish abundance, after removing one outlier of concern with a comparatively large sample size relative to the other effect sizes (k=76). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). Diamond: overall mean effect size of random-effects model.

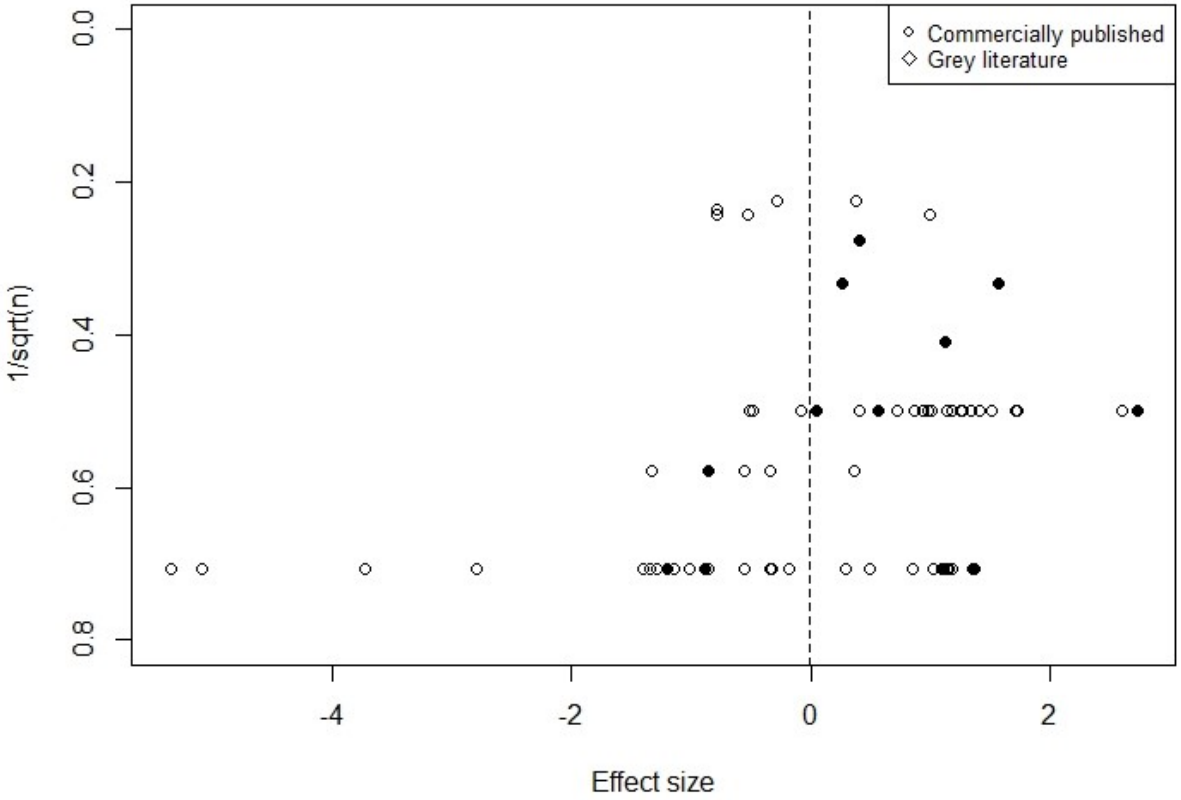

Fig. S5 Funnel plot showing *Control/Impact* studies of abundance (k = 76), with the one influential outlier removed. Open circles indicated datasets from commercially published articles and filled circles indicate datasets from grey literature. Summary effect is indicated by the dashed line.

Biomass

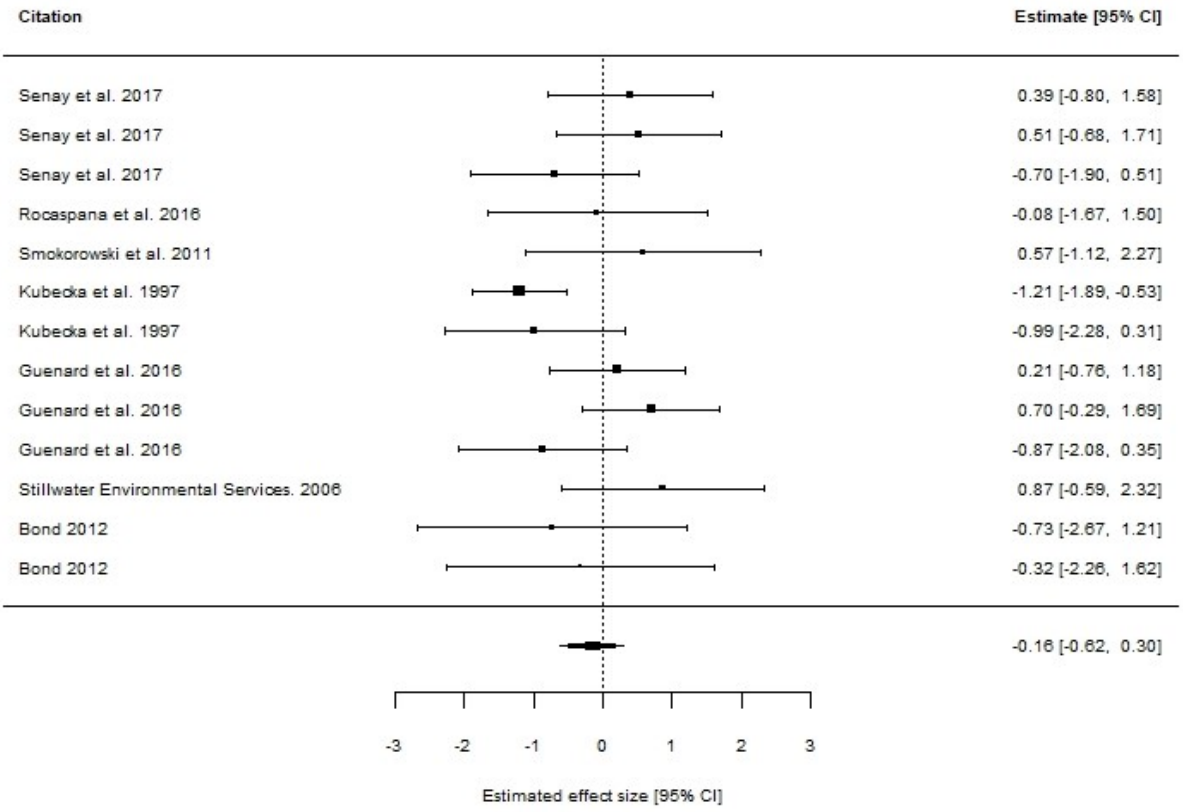

Fig. S6. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on fish biomass (k=13). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). Diamond: overall mean effect size of random-effects model.

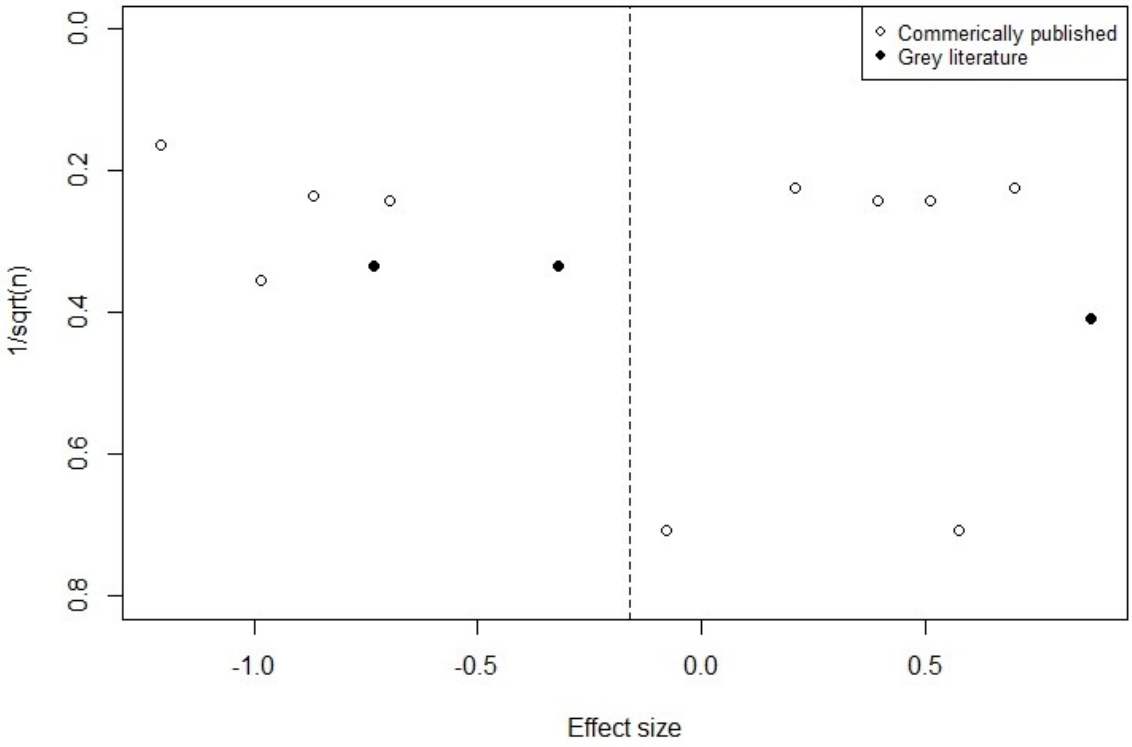

Fig. S7. Funnel plot showing *Control/Impact* studies of biomass (k = 13). Open circles indicate datasets from commercially published articles and filled circles indicate datasets from unpublished literature. Summary effect is indicated by the dashed line.

Moderator analysis – *Control/Impact* Studies

Abundance

Table S1. Pearson chi-squared values (above diagonals), and their *p*-values (below diagonals) of contingency analysis for independence of moderators* considered for *CI* study designs and abundance.

Moderator(s)	Dam size	Hydropower operational regime	Direction of flow magnitude alteration	Alterations to other flow components	Sampling method	Sampling season	Type of comparator (spatial)	Time since intervention	Monitoring duration
Dam size	-	29.22 (77)	1.92 (76)	30.83 (77)	47.35 (73)	65.82 (77)	3.7 (76)	29.95 (77)	53.46 (75)
Hydropower operational regime	<0.0001	-	45.75 (76)	11.36 (77)	2.94 (73)	40.81 (77)	41.78 (76)	82.56 (77)	53.28 (75)
Direction of flow magnitude alteration	0.382	<0.0001	-	6.33 (76)	0 (72)	10.31 (76)	43.89 (75)	52.25 (76)	29.38 (74)
Alterations to other flow components	<0.0001	0.023	0.042	-	38.23 (73)	31.62 (77)	12.52 (76)	10.92 (76)	314.32 (75)
Sampling method	<0.0001	0.230	1.000	<0.0001	-	41.27 (73)	8.92 (73)	10.01 (73)	34.99 (71)
Sampling season	<0.0001	<0.0001	0.016	<0.0001	<0.0001	-	0.06 (76)	24.18 (77)	36.75 (75)
Type of comparator (spatial)	0.157	<0.0001	<0.0001	0.002	0.004	0.996	-	65.06 (76)	23.17 (75)
Time since intervention	<0.0001	<0.0001	<0.0001	0.028	0.007	0.0005	<0.0001	-	40.32 (75)
Monitoring duration	<0.0001	<0.0001	<0.0001	0.006	<0.0001	<0.0001	<0.0001	<0.0001	-

*(i) Intervention related moderators: dam size, hydropower operational regime, direction of flow magnitude alteration; (ii) confounder related moderators: alterations to other flow components, time since intervention; (iii) study design related moderators: sampling method, sampling season, type of comparator (spatial), monitoring duration.

Global meta-analysis – Within-year *Before/After* studies
Abundance

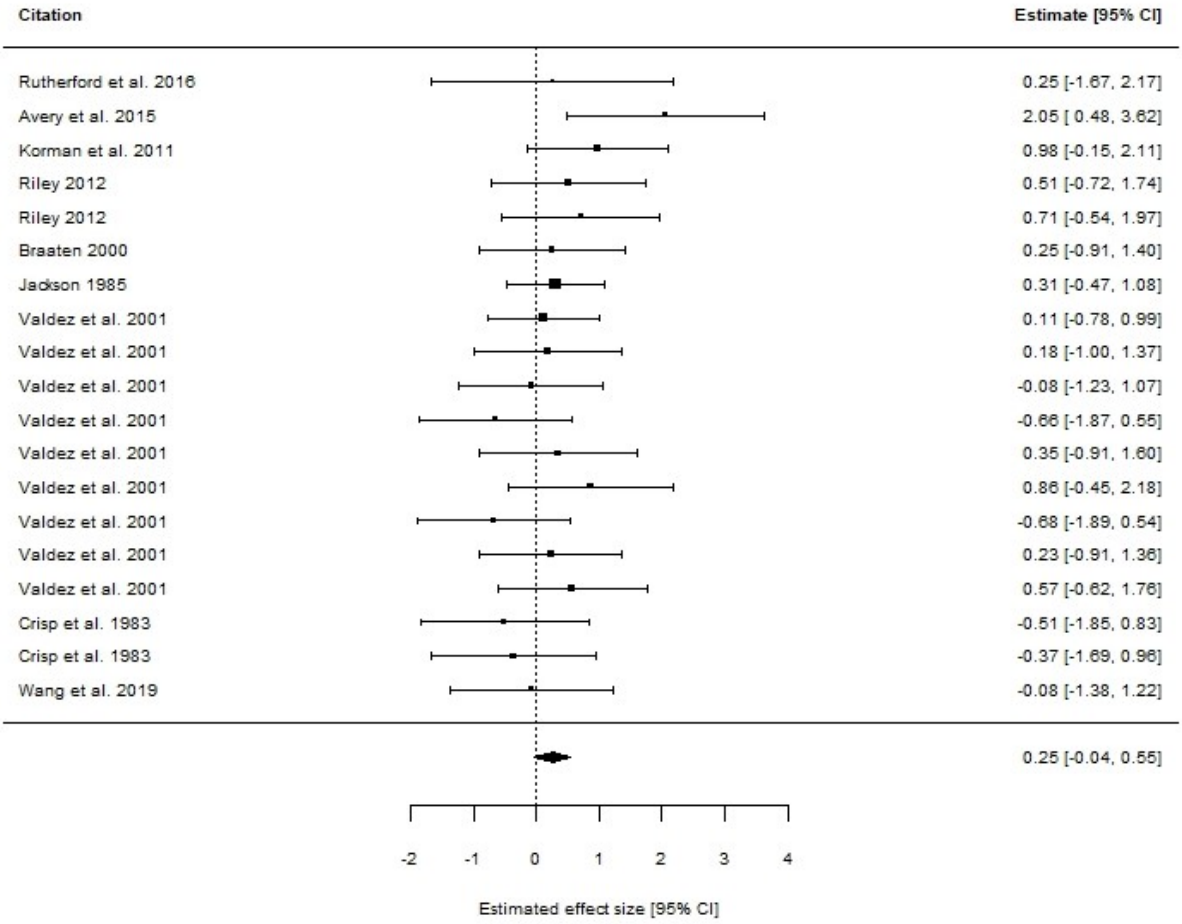

Fig. S8. Summary plot of all effect size estimates from within-year *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance after year 1 (k=19). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

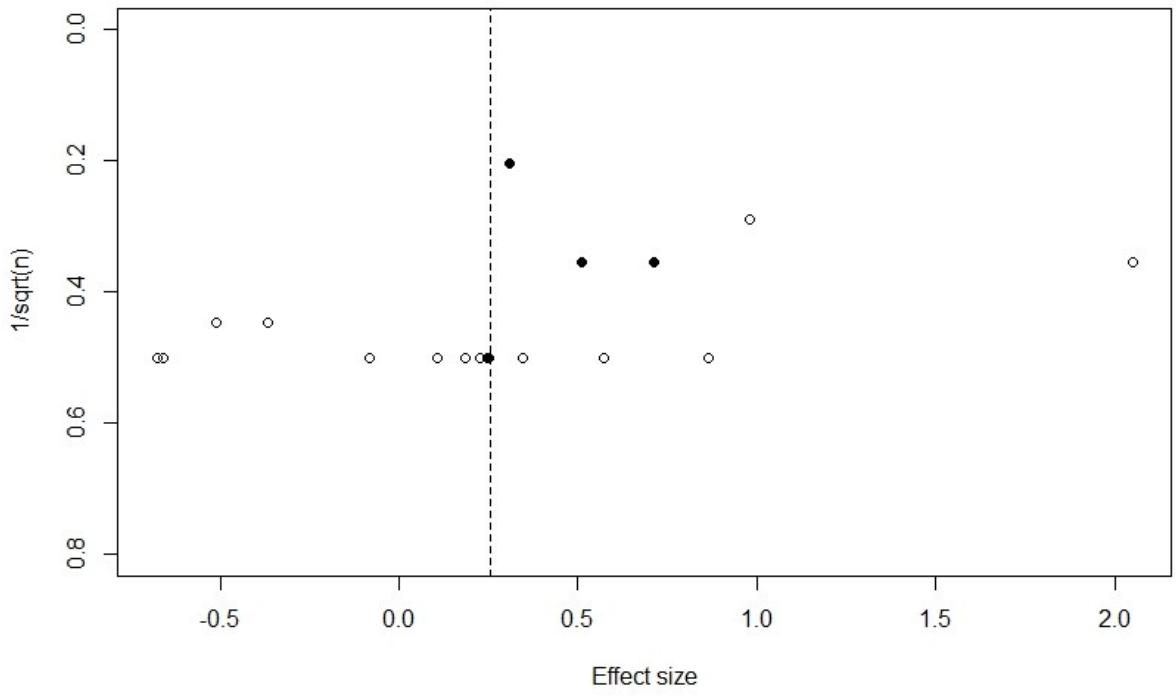

Fig. S9. Funnel plot showing publication bias for within- year *Before/After* studies of abundance (k = 19). Open circles indicate datasets from commercially published articles and filled circles indicate datasets from grey literature. Summary effect is indicated by dashed line.

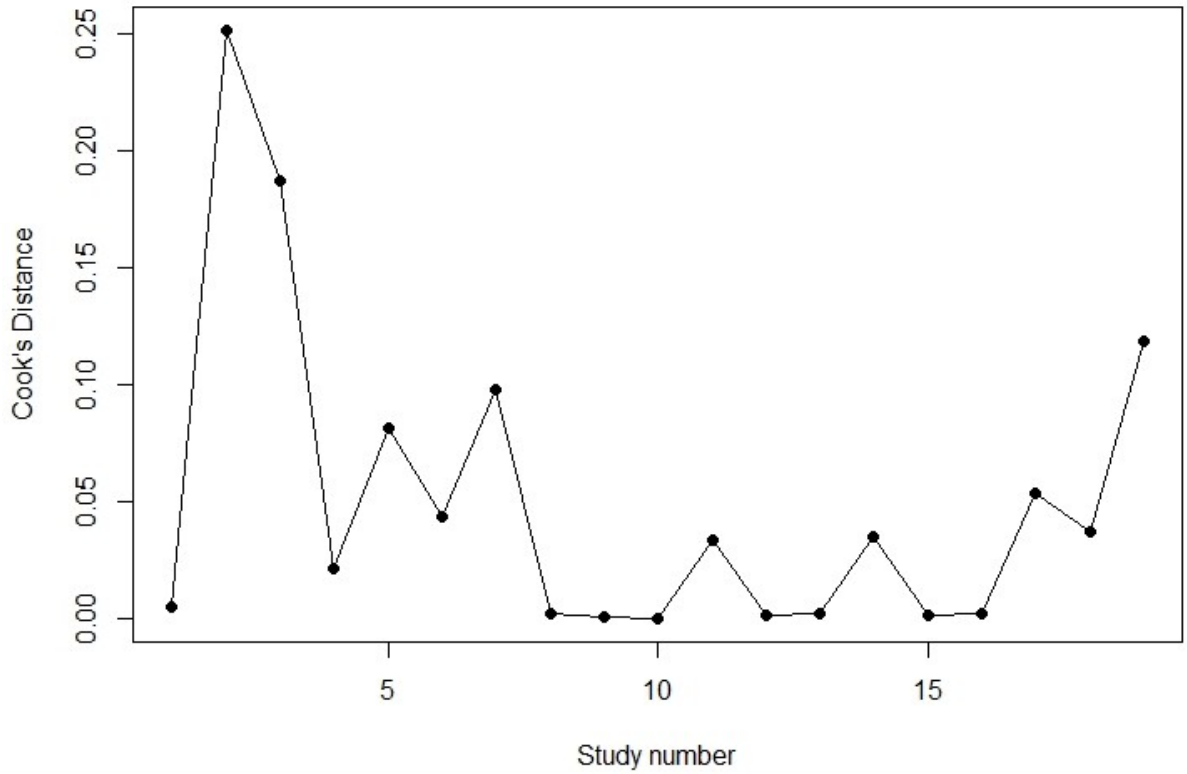

Fig. S10. Cook's distance plot indicating influence of effect sizes.

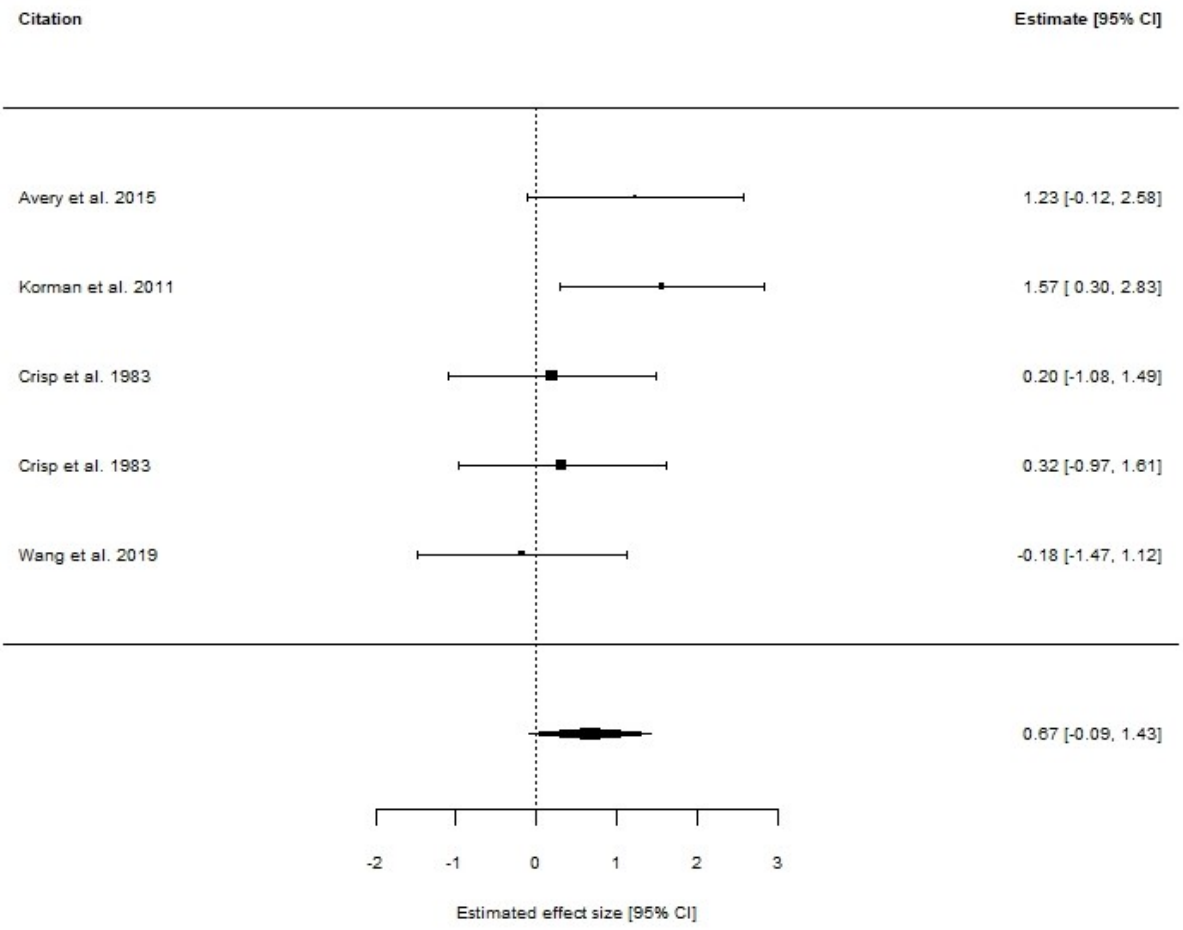

Fig. S11. Summary plot of all effect size estimates from within-year *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance after year 2 (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

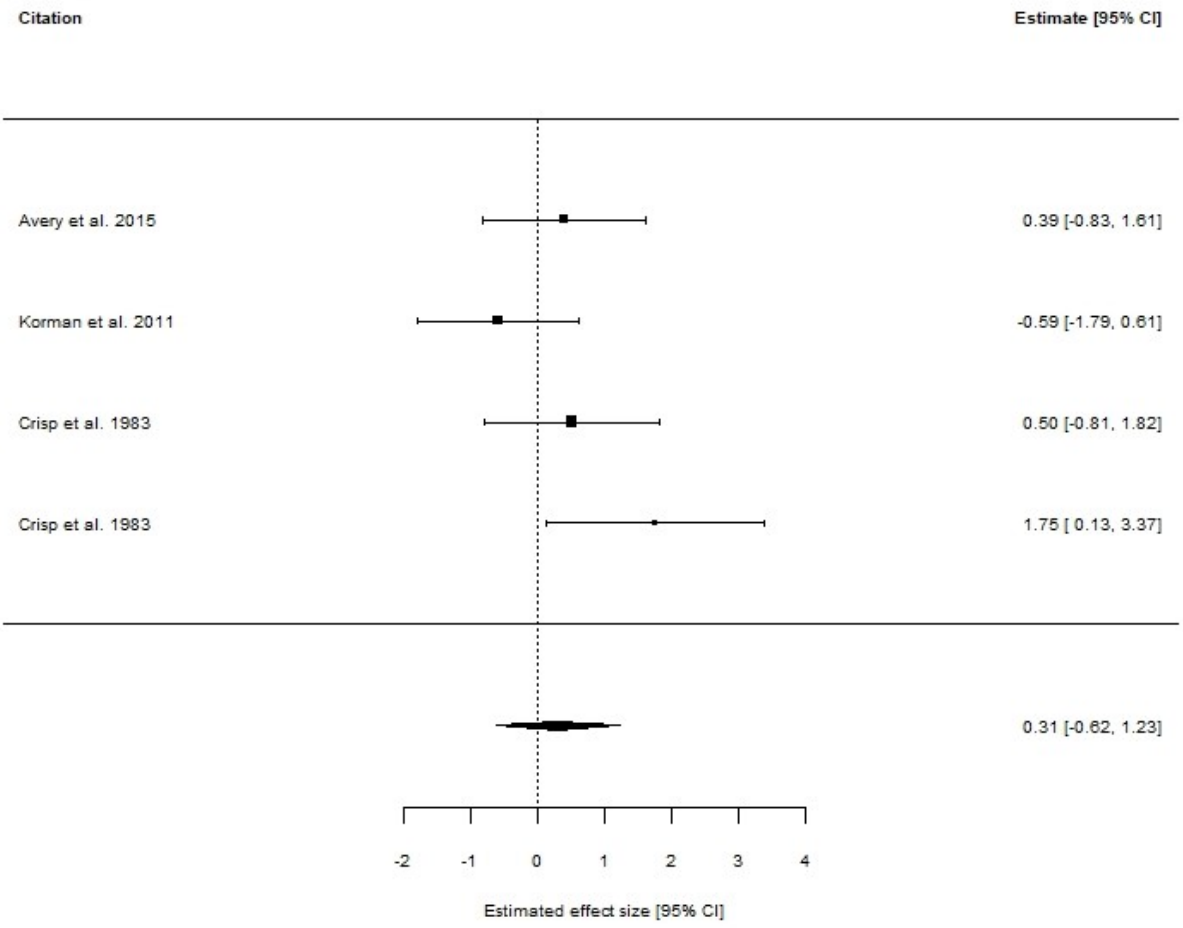

Fig. S12. Summary plot of all effect size estimates from Within-year *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance after year 3 (k=4). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

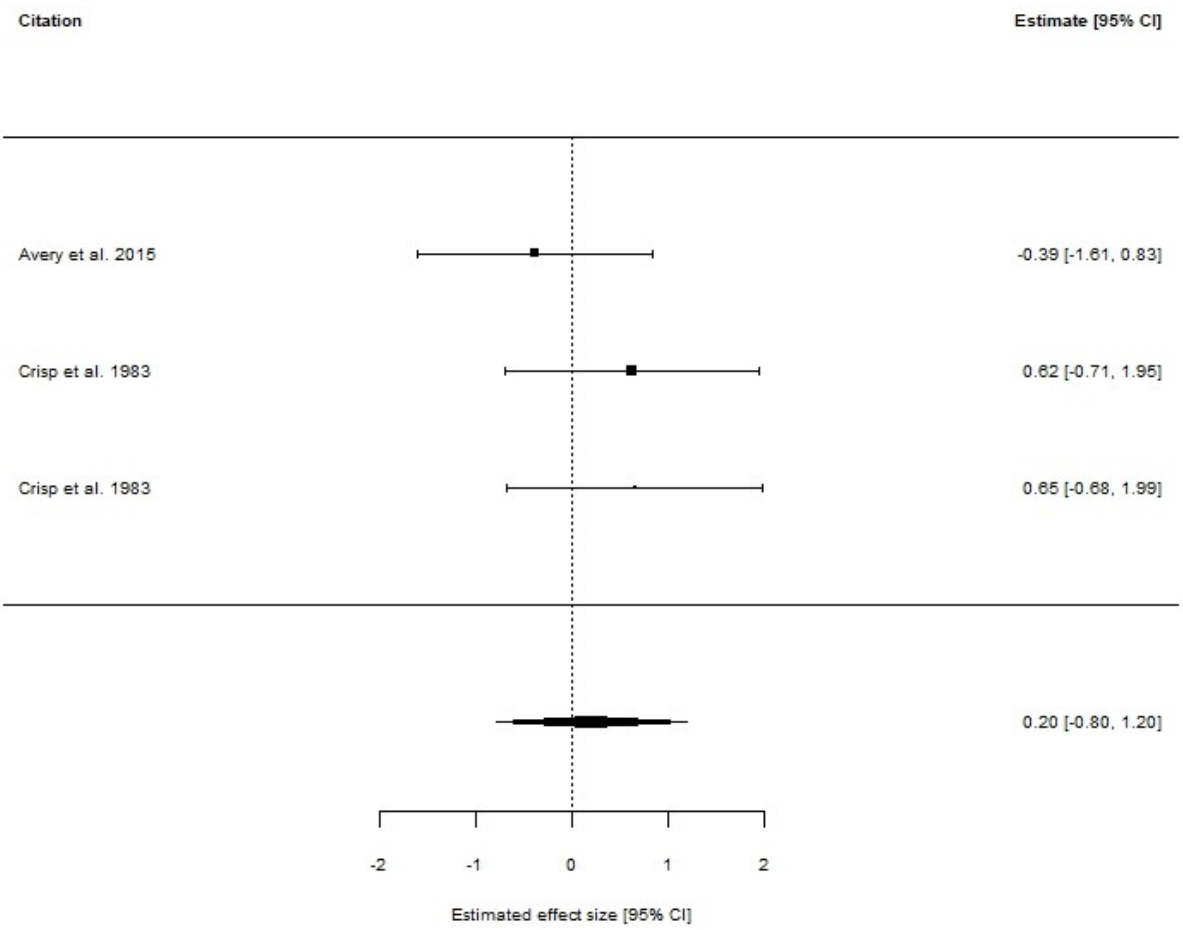

Fig. S13. Summary plot of all effect size estimates from Within-year *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance after year 4 (*k*=3). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

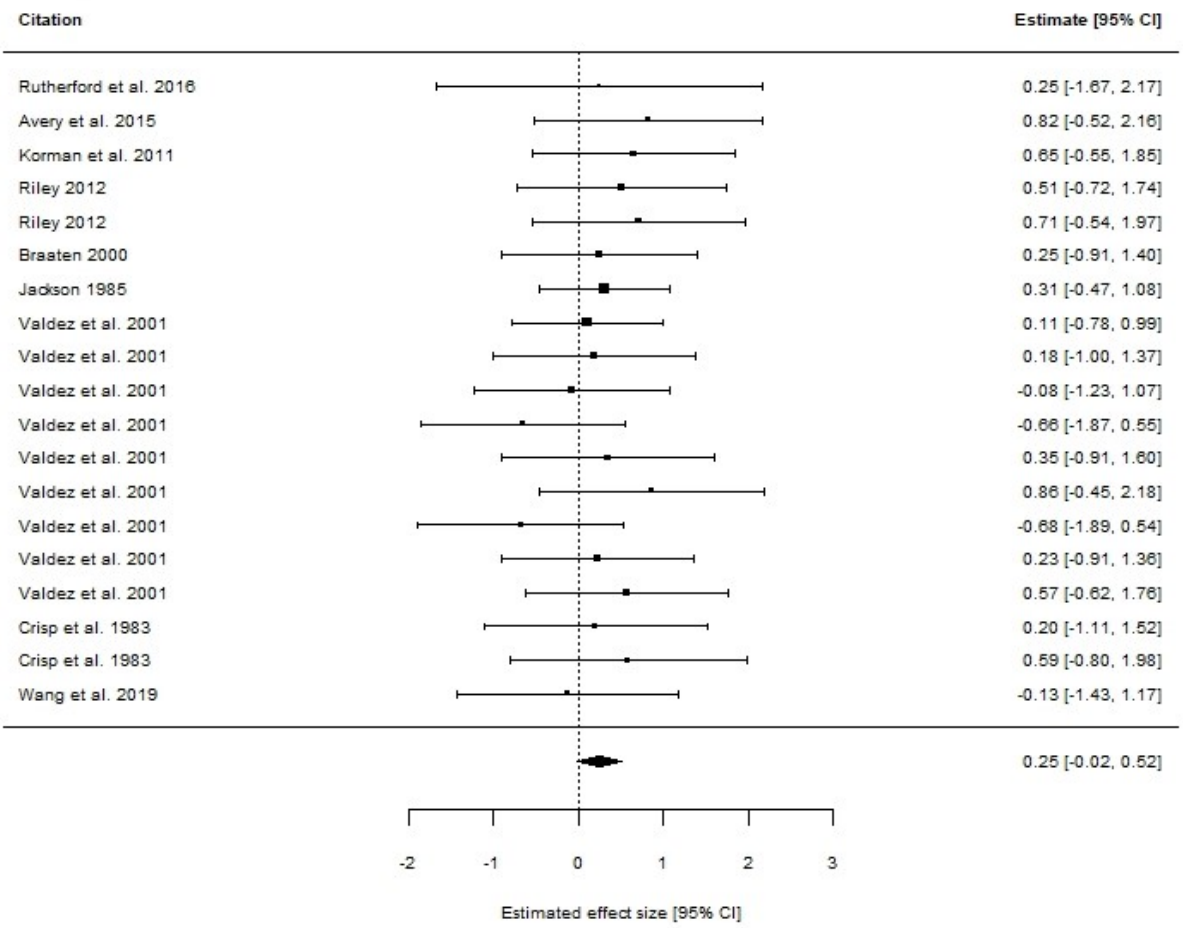

Fig. S14. Summary plot of all effect size estimates from Within-year *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance after year 1- 4 aggregated (*k*=19). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Sensitivity analysis (Abundance, After year 1)

For this analysis, to test the influence of studies where stocking potentially occurred and flow magnitude components are not specified, the same two datasets (which both had these features) were removed and the average effect size for fish abundance was determined with the remaining studies.

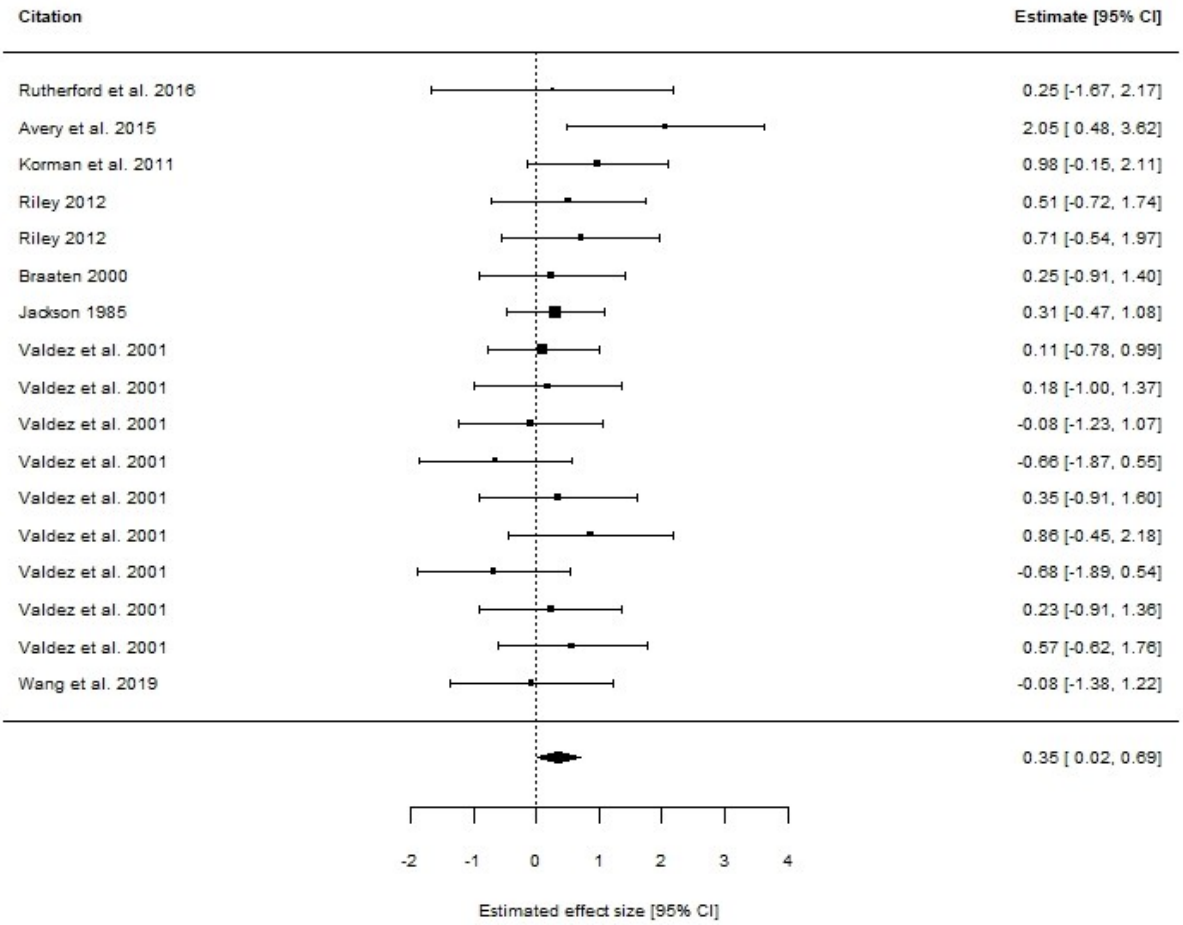

Fig. S15. Summary plot of all effect size estimates from within-year *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance after year 1 (k=17), considering only studies where it is clear that no stocking influenced the waterbody during sampling, and flow magnitude components are stated. Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Global meta-analysis – Interannual *Before/After* studies
Abundance

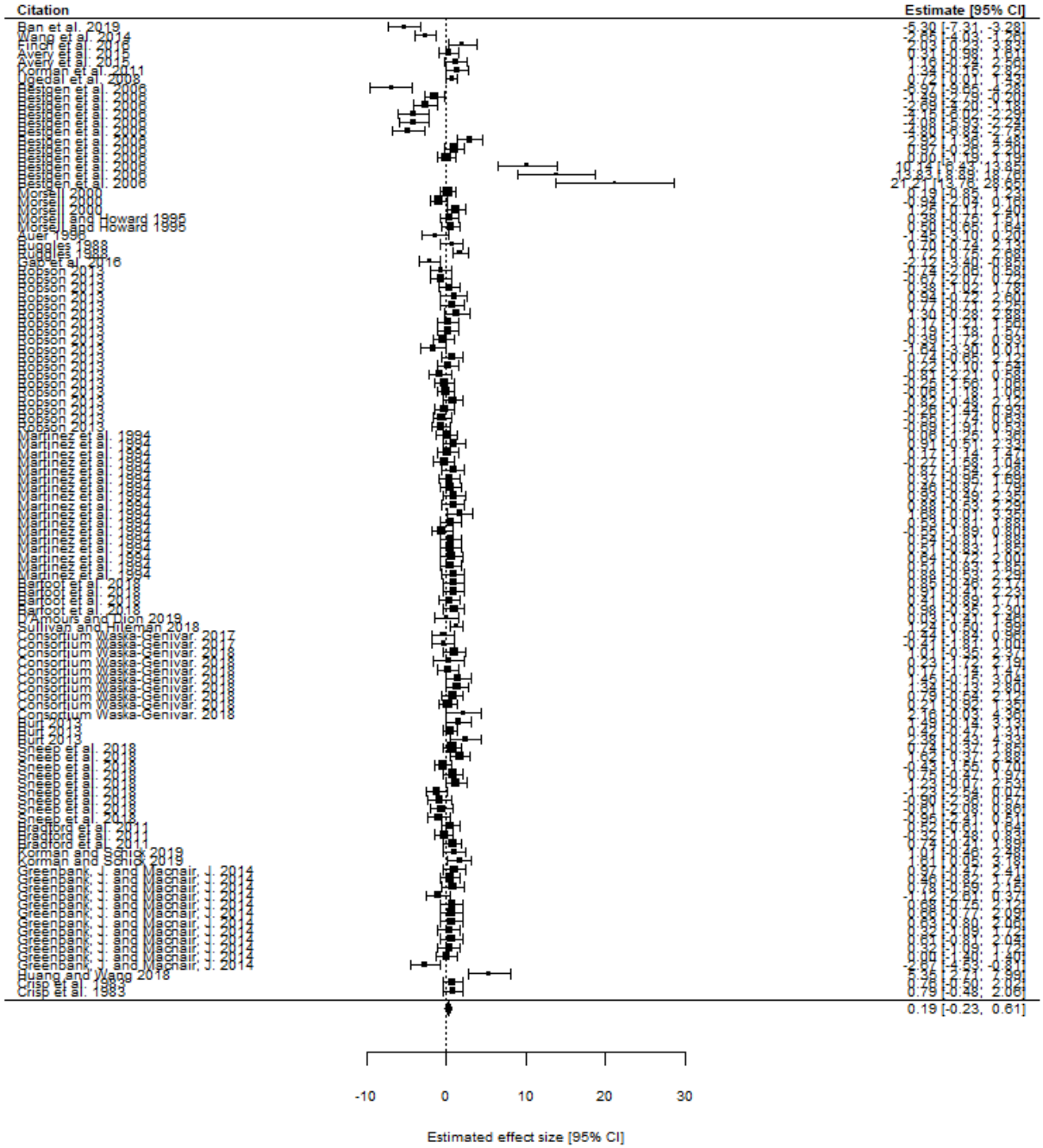

Fig. S16. Summary plot of all effect size estimates from interannual *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance (k=112). Datasets from the same study with different interventions were compared to a single comparator. Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

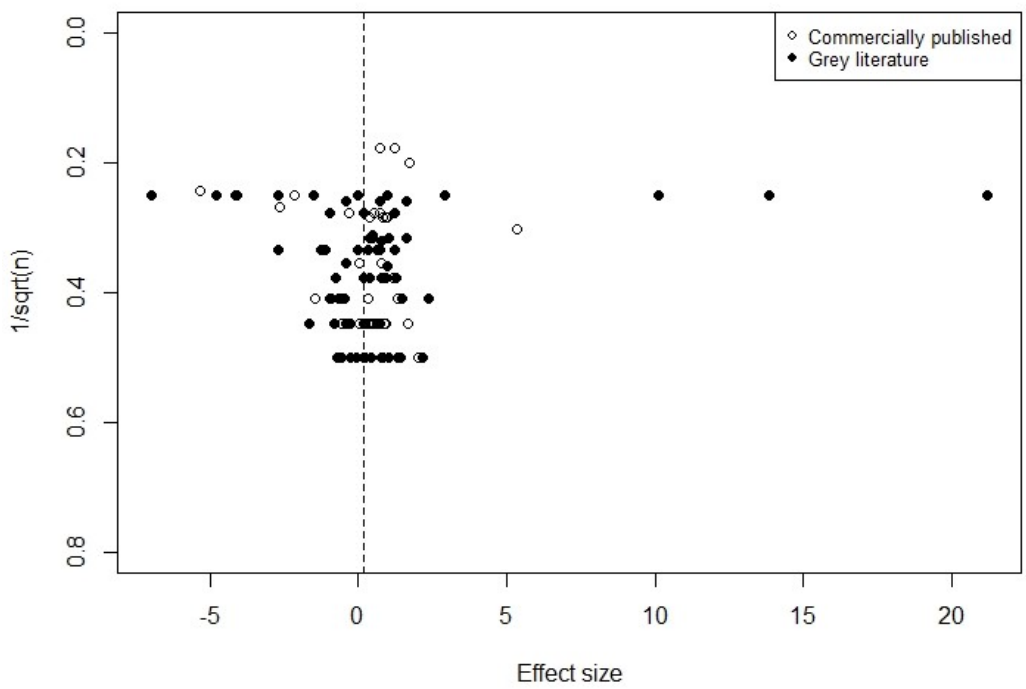

Fig. S17. Funnel plot showing interannual *Before/After* studies of abundance (k = 112). Open circles indicate datasets from commercially published articles and filled circles indicate datasets from grey literature. Summary effect is indicated by the dashed line.

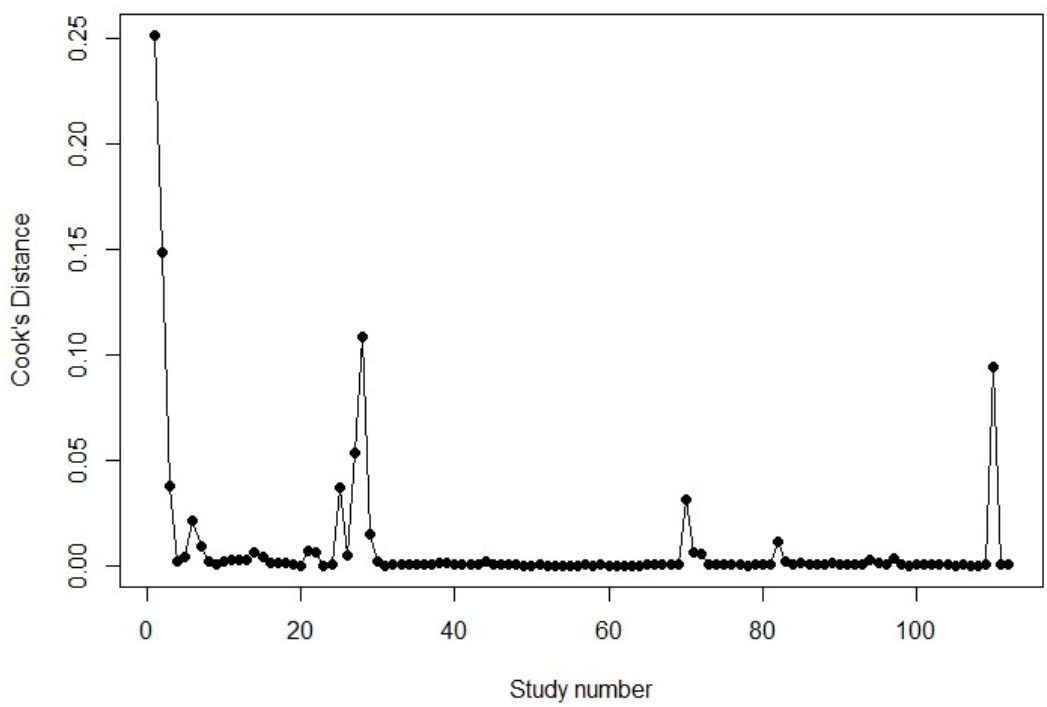

Fig. S18. Cook's distance plot indicating influence of effect size. Note the outlier of concern (Cook's distance ≈ 0.25).

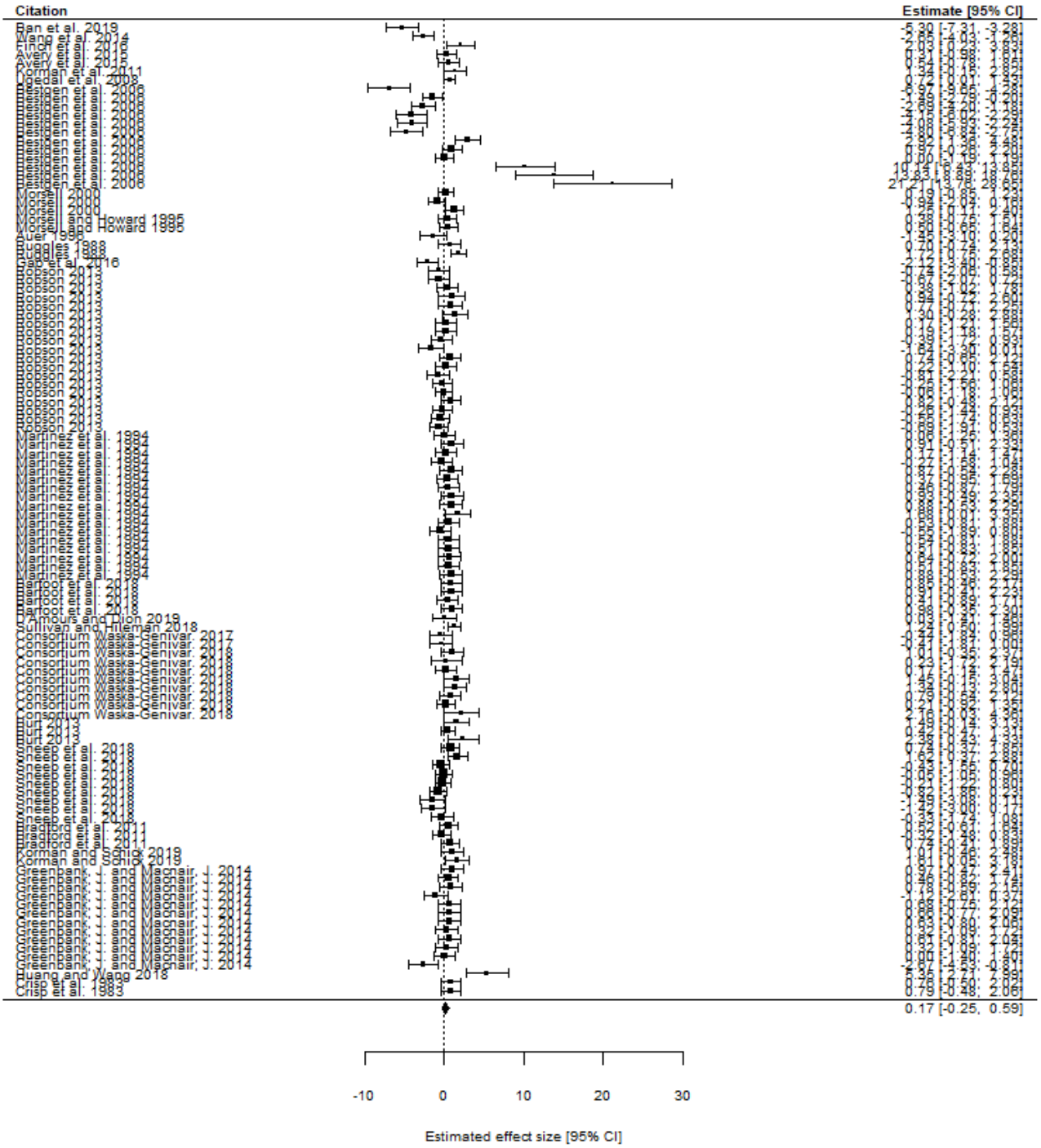

Fig. S19. Summary plot of all effect size estimates from interannual *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance (*k*=112). Datasets from the same study with different interventions were compared to each previous period (i.e., Trial 2 was compared to Trial 1, rather than the original pre-trial period). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model. Diamond: overall mean effect size of random-effects model.

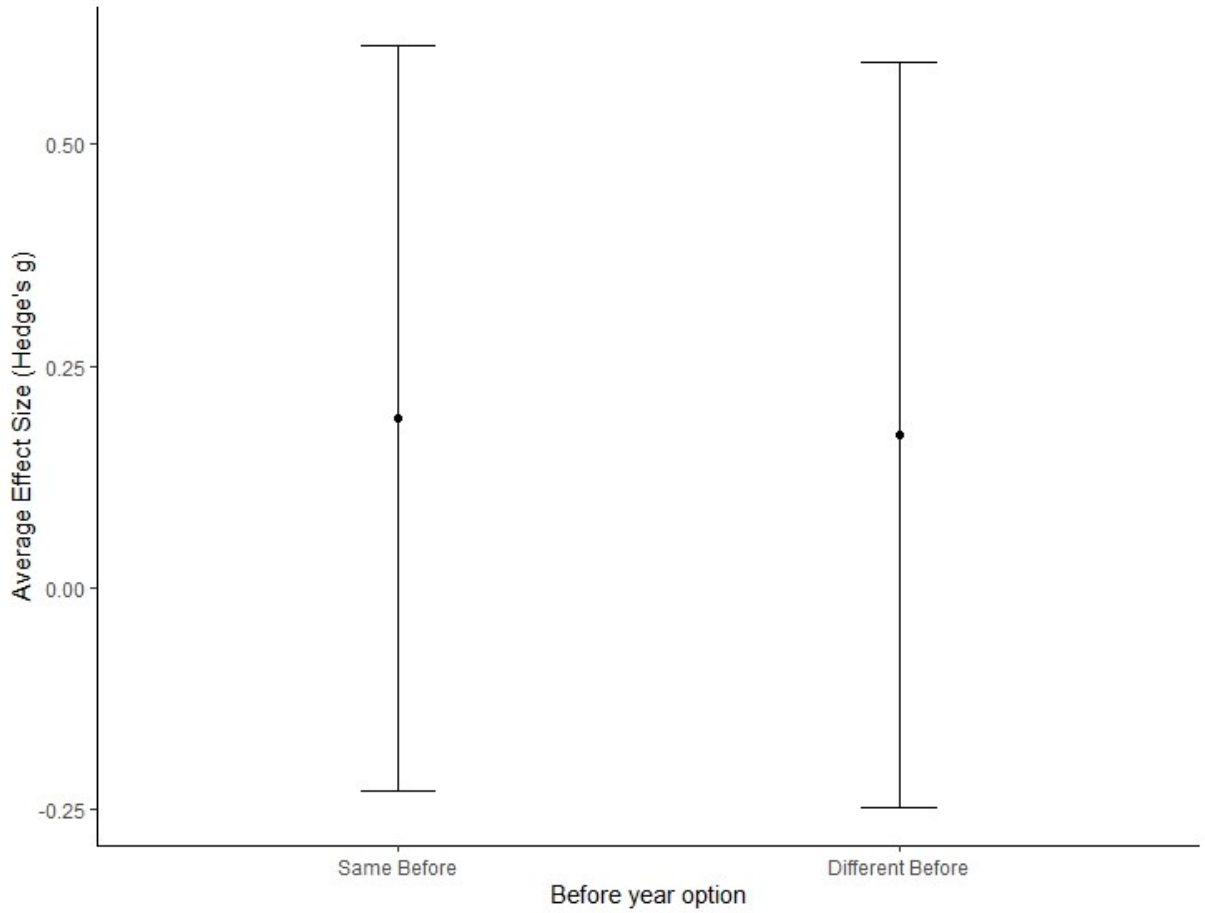

Fig. S20. Comparison of interannual *Before/After* studies' overall effect size when datasets from studies with more than one intervention are compared to a single pre-trial *Before* period (Same *Before*) or when each subsequent intervention is compared to the previous period (i.e., Trial 2 is compared to Trial 1, rather than to the pre-trial period).

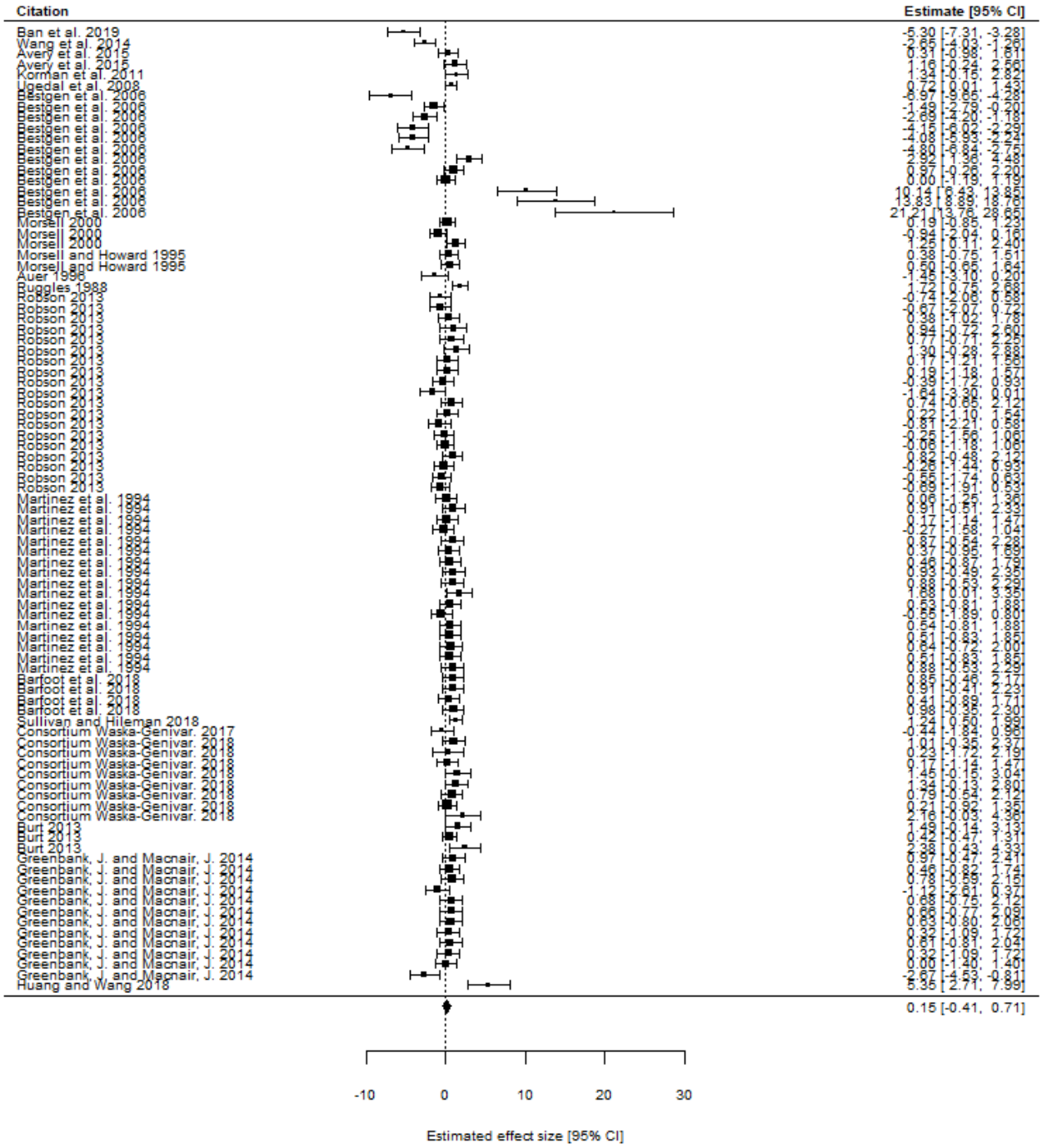

Fig. S21. Summary plot of all effect size estimates from interannual *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance that reported fish outcomes as either single datapoints per year, or sums per year which were then averaged in the *Before* and *After* periods (k=91). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

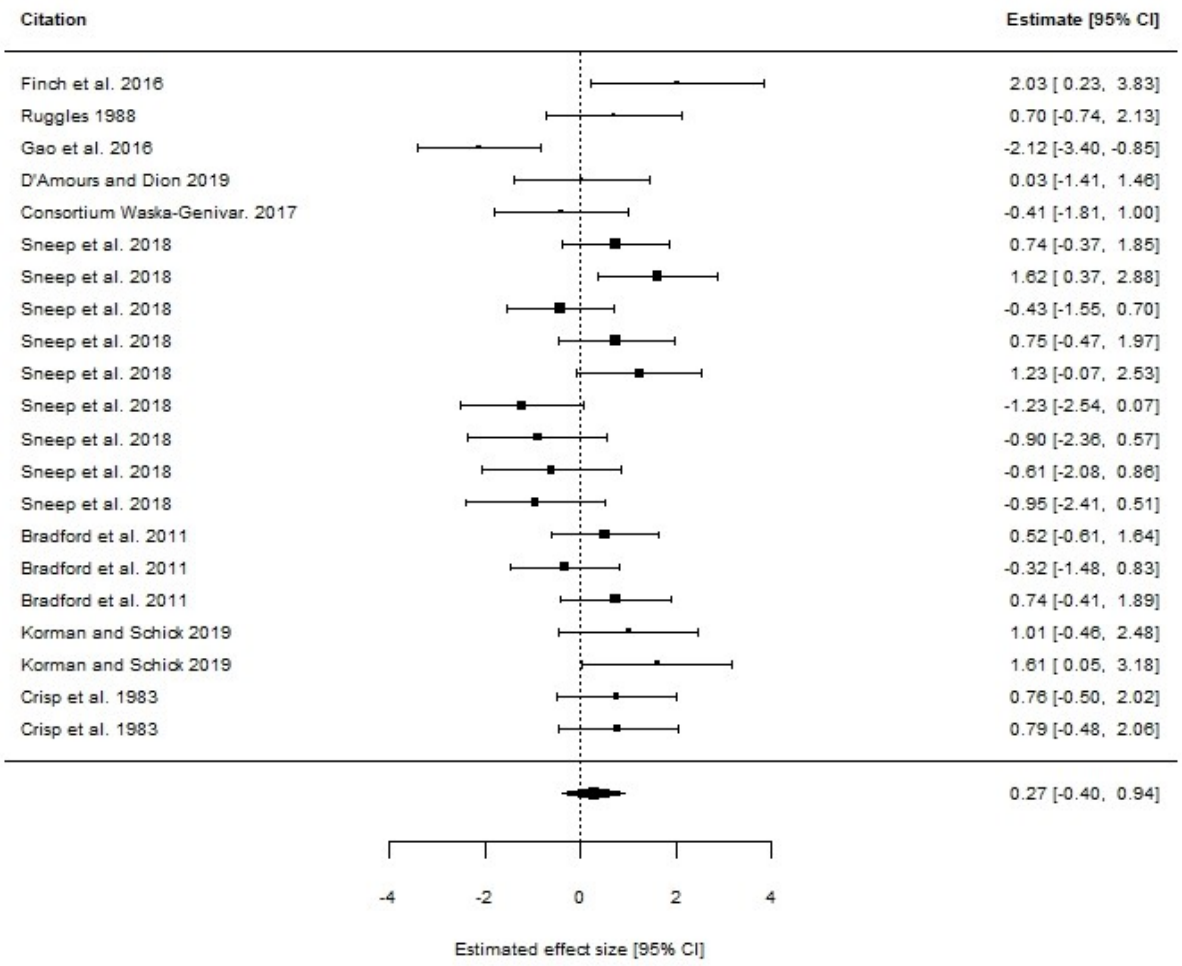

Fig. S22. Summary plot of all effect size estimates from interannual *Before/After* evaluations of the impact of flow magnitude alterations on fish abundance that reported fish outcomes as averages of averages for the *Before* and *After* periods (k=21). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

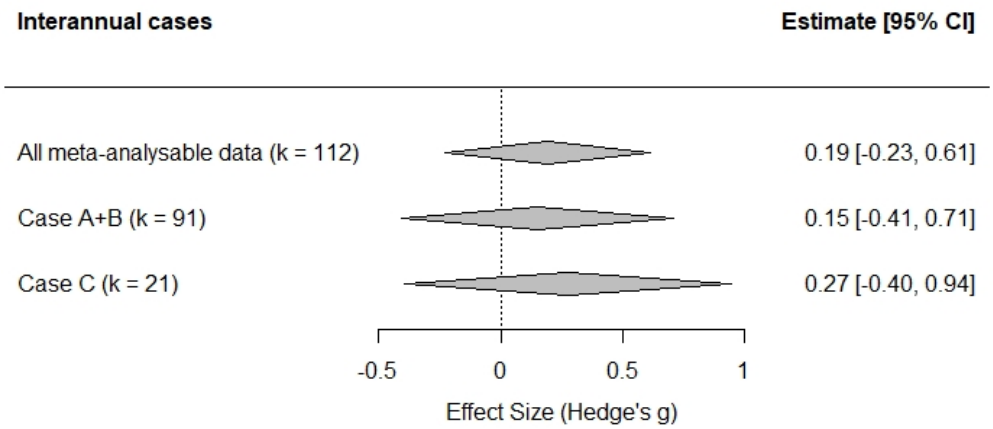

Fig. S23. Comparison of effect sizes for three types of interannual *Before/After* data, for the impact of flow magnitude alterations on fish abundance. *Case A*: fish sampled only once per year; *Case B*: studies only report total fish abundance from multiple sampling seasons within a given year; *Case C*: fish abundances sampled/reported more than once per year, averaged per year and then averaged across all *Before* year and all *After* years (averages of averages). Cases are compared to all meta-analysable data, which includes all cases of *A*, *B* and *C*.

Biomass

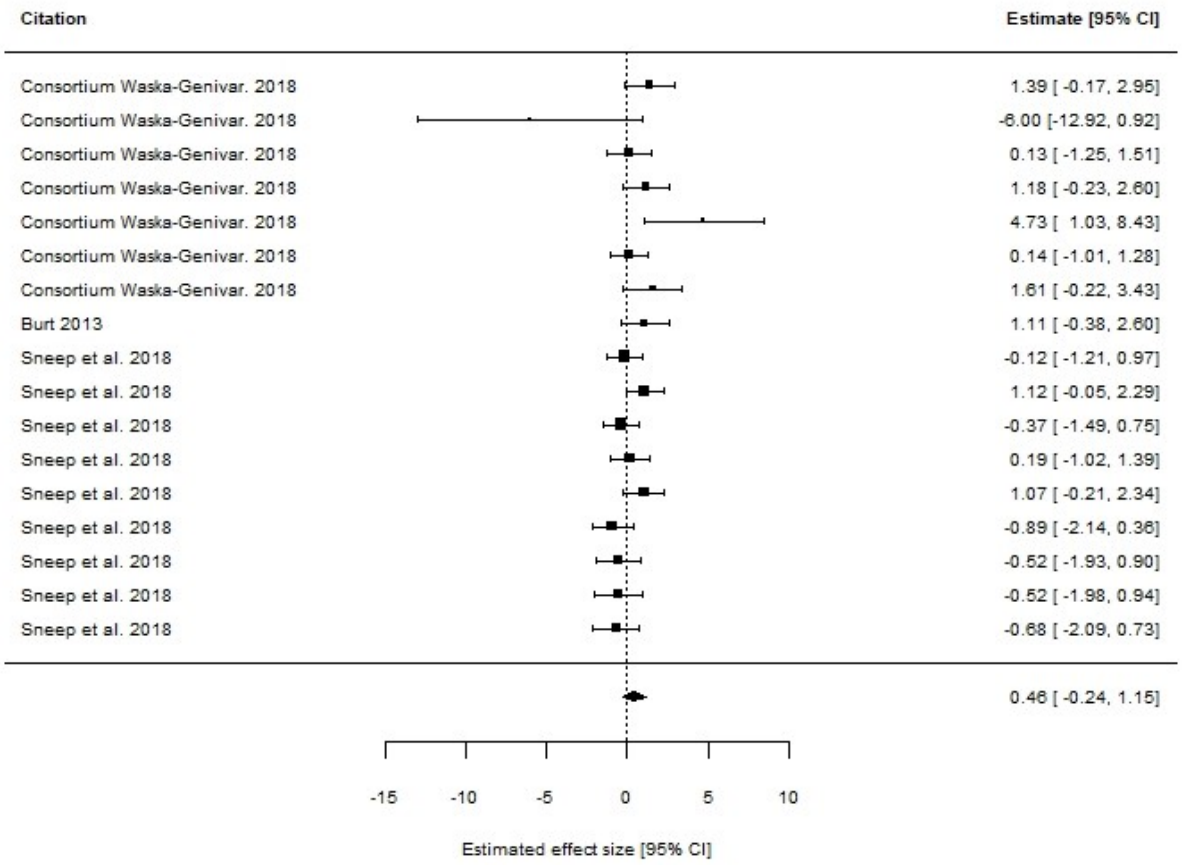

Fig. S24. Summary plot of all effect size estimates from interannual *Before/After* evaluations of the impact of flow magnitude alterations on fish biomass (*k*=17). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the biomass was higher in the *After* period than in the *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

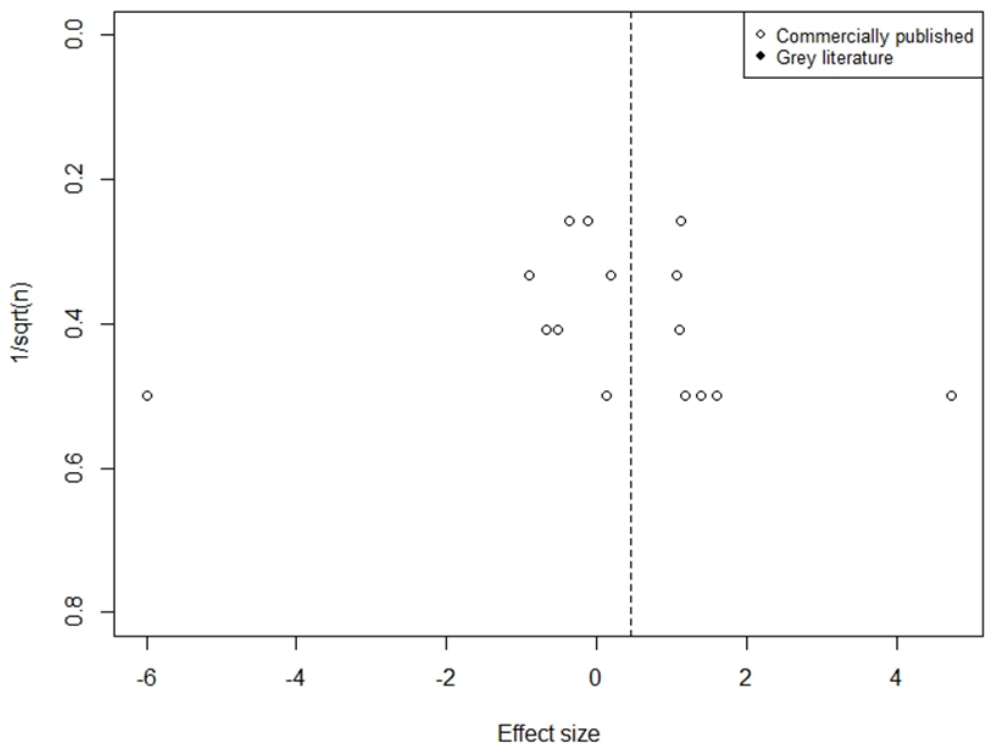

Fig. S25. Funnel plot showing *Control/Impact* studies of abundance (k = 77). Open circles indicate datasets from commercially published articles and filled circles indicate datasets from grey literature. Summary effect is indicated by the dashed line.

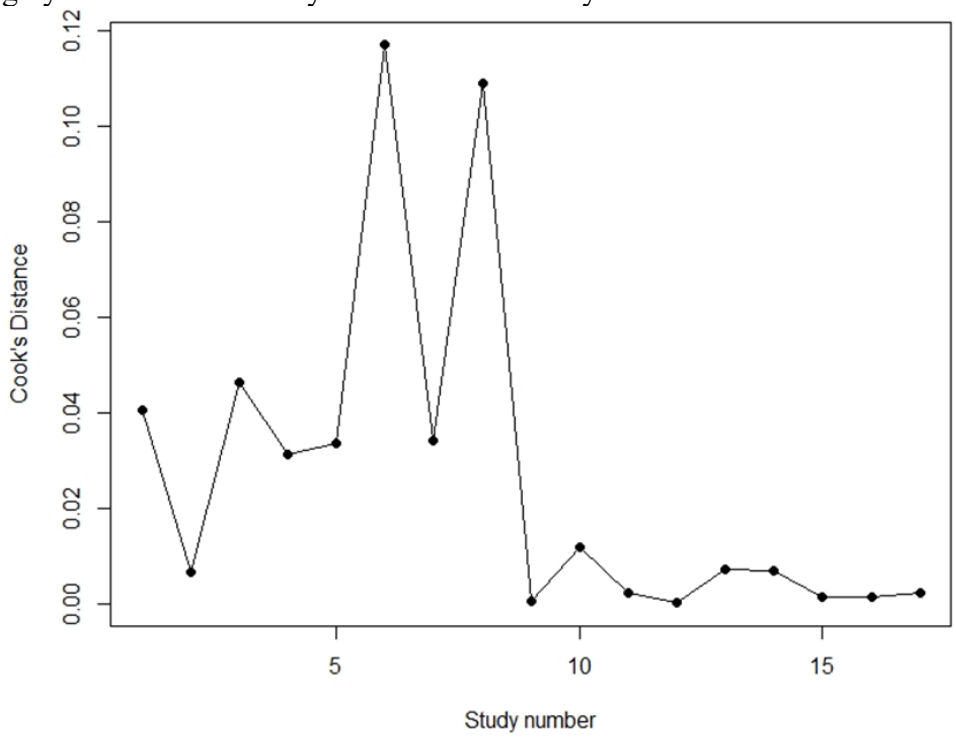

Fig. S26. Cook's distance plot indicating influence of effect size for biomass interannual *Before/After* studies.

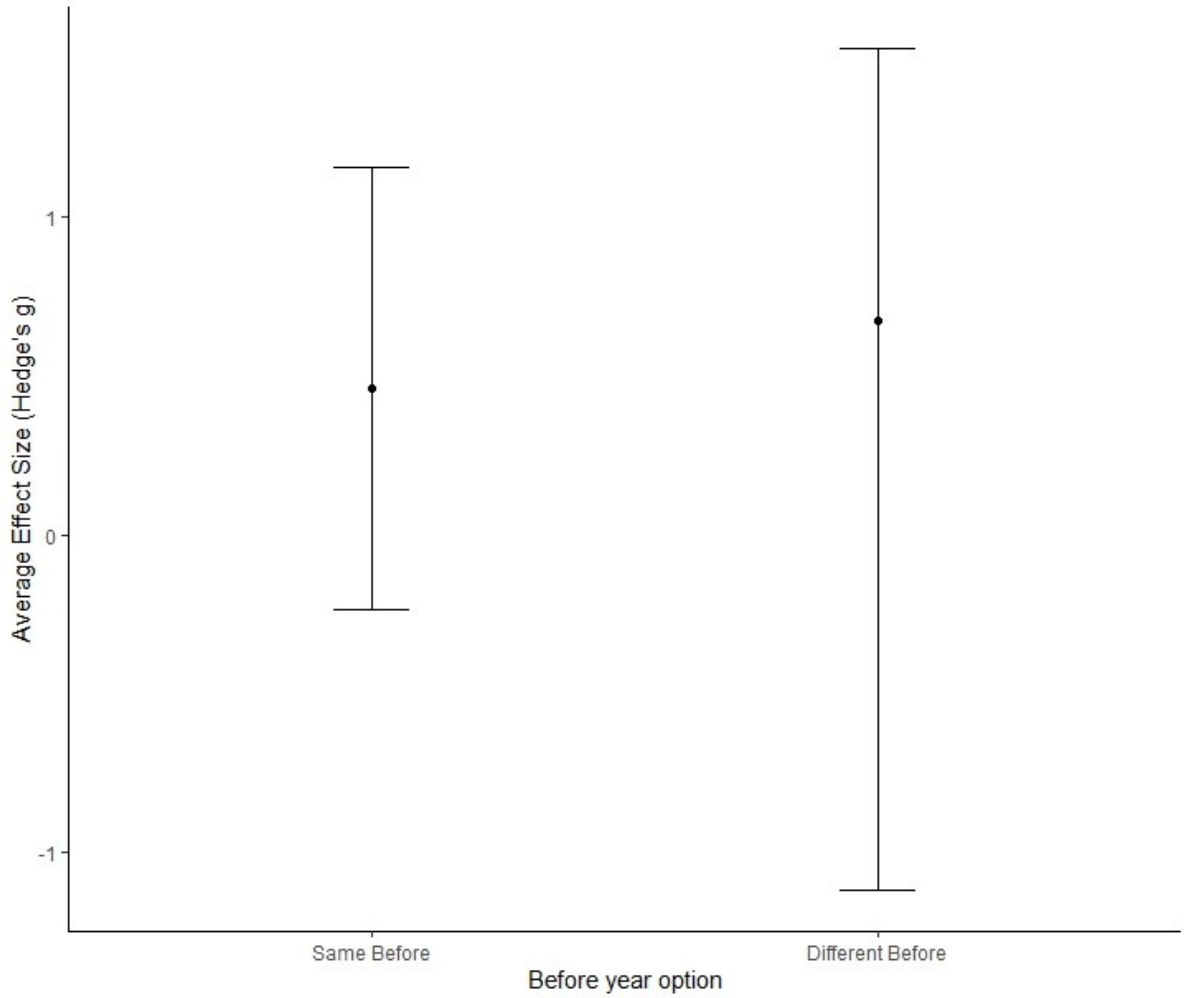

Fig. S27. Comparison of interannual *Before/After* studies overall biomass effect size when datasets from studies with more than one intervention are compared to a single pre-trial *Before* period (Same *Before*) or when each subsequent intervention is compared to the previous period (i.e., Trial 2 is compared to Trial 1, rather than to the pre-trial period).

Moderator analysis – Interannual *Before/After* Studies

Abundance

Table S2. Pearson chi-squared values (above diagonals), and their *p*-values (below diagonals) of contingency analysis for independence of moderators* considered for *BA* study designs and abundance.

Moderator(s)								
	Hydropower Operational regime	Direction of flow magnitude change	Alterations to other flow components	Sampling method	Sampling season	Type of comparator (temporal)	Time since intervention	Life stage
Hydropower Operational regime	-	68.32 (112)	78.52 (112)	13.65 (112)	93.14 (112)	42.18 (112)	73.11 (111)	61.33 (108)
Direction of flow magnitude change	<0.0001	-	121.47 (112)	15.03 (112)	53.68 (112)	50.3 (112)	38.07 (111)	45.56 (108)
Alterations to other flow components	<0.0001	<0.0001	-	11.5 (112)	46.34 (112)	69.16 (112)	44.36 (111)	51.26 (108)
Sampling method	0.136	0.090	0.243	-	65.89 (112)	9.69 (112)	9.28 (111)	30.44 (108)
Sample season	<0.0001	<0.0001	<0.0001	<0.0001	-	14.09 (112)	89.84 (111)	54.04 (108)
Type of comparator (temporal)	<0.0001	<0.0001	<0.0001	0.021	0.007	-	25.24 (111)	22.97 (108)
Time since intervention	<0.0001	<0.0001	<0.0001	0.158	<0.0001	<0.0001	-	24.18 (107)
Life stage	<0.0001	<0.0001	<0.0001	0.002	<0.0001	0.0001	0.002	-

*(i) Intervention related moderators: hydropower operational regime, direction of flow magnitude alteration; (ii) confounder related moderators: alterations to other flow components, time since intervention, life stage; (iii) study design related moderators: sampling method, sampling season, type of comparator (spatial), monitoring duration.

Interannual BA: Abundance - Meta-regression

Due to two extreme outliers in effect sizes (*Ictalurus punctatus* and *Micropterus dolomieu*; (Bestgen et al. 2006), we were unable to achieve normality through transformation for the continuous moderator 'monitoring duration'. We therefore conducted meta-regression with and without these outliers and present results for both analyses below (Fig. S28 and S29). We found no significant relationship of fish abundance and monitoring duration in either instance and the results of the two models did not differ greatly. We report results of Q_M for the models in Table S3. Potential reasons for the occurrence of these outliers for *Ictalurus punctatus* (ES = 14) and *Micropterus dolomieu* (ES = 21), may be due to factors other than flow magnitude alterations that occurred in the *Before* and *After* periods of the study (Bestgen et al. 2006). During the *Before* sampling period, *I. punctatus* had the lowest numbers ever recorded in the history of sampling in the system; this results in a comparison between the *After* period and a non-representative *Before* period which might inflate the effect size. *Micropterus dolomieu* is an established invader with an active removal program, conducted throughout the *Before* and *After* periods. This species had a very successful spawning year during the *After* sampling period, probably due to warm water temperatures, which led to an extreme number of age-1 fish (Bestgen et al. 2006). The *Before* period may have had a depressed number of fish due to active removal and the *After* period had an unexpectedly high number of age-1 fish, resulting in an inflated effect size.

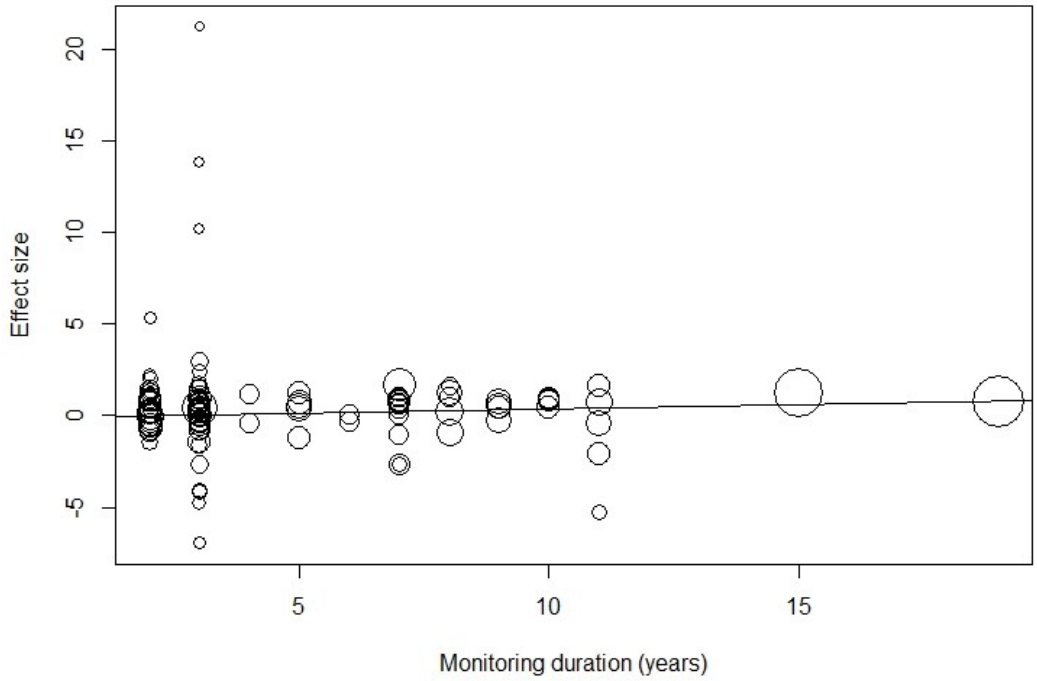

Fig. S28. Meta-regression of effect sizes (Hedge's *g*) against monitoring duration (years) for interannual *Before/After* studies and abundance. Two extreme effect sizes were retained in this model to compare to model without outliers.

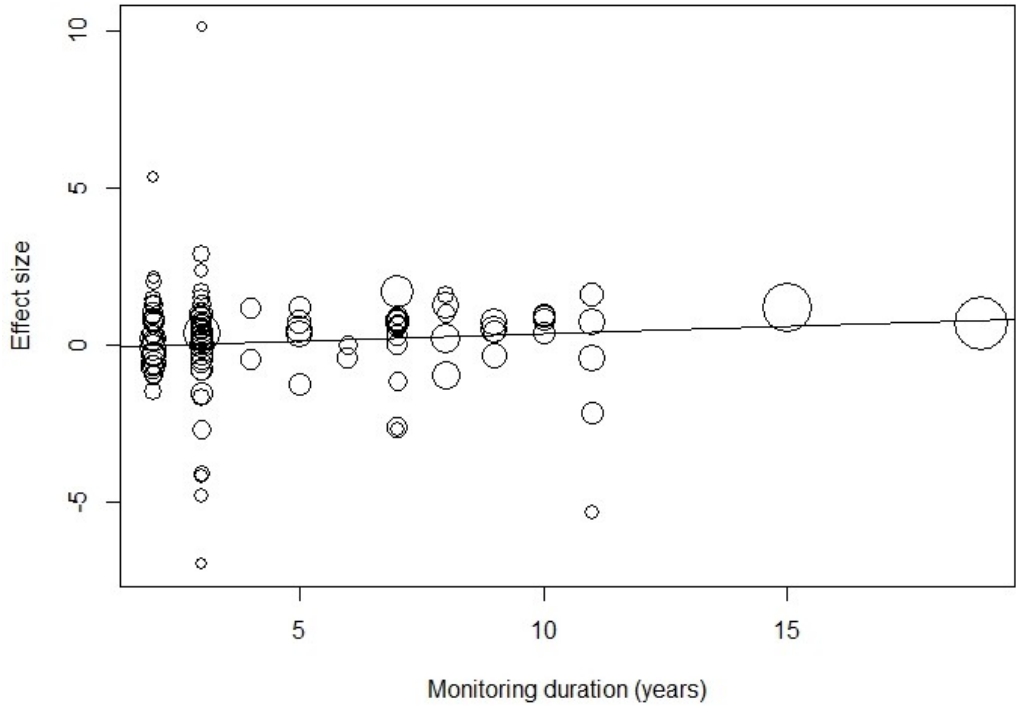

Fig. S29 Meta-regression of effect sizes (Hedge's *g*) against monitoring duration (years) for interannual *Before/After* studies and abundance. Two extreme effect sizes were removed to improve model fit.

Table S3. Summary results of meta-regression using subsets of fish abundance effect sizes for interannual *Before/After* studies, testing the influence of monitoring duration with and without extreme outliers.

Moderator	***k***	***Q* statistic (p-value)**	**Q_M (*p*-value)**	**Q_E (*p*-value)**
Monitoring duration (with outliers)	112	**421.12 (*p*<0.0001)**	-	-
Unmoderated model	112	-	1.32 (*p*=0.252)	**417.43 (*p*<0.0001)**
Monitoring duration				
Monitoring duration (without outliers)				
Unmoderated model	110	**361.71 (*p*<0.0001)**	-	-
Monitoring duration	110	-	1.39 (*p*=0.239)	**357.43 (*p*<0.0001)**

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. Significance at $p < 0.05$; * Significance at $p < 0.1$.

Appendix 12. Taxonomic analysis

Description: Includes forest plots for all families with sufficient sample sizes and for, families with significant heterogeneity, genera therein with sufficient sample size for further analysis (i.e., ≥3 datasets from ≥2 independent studies.

***Control/Impact* Studies: Family**

Abundance

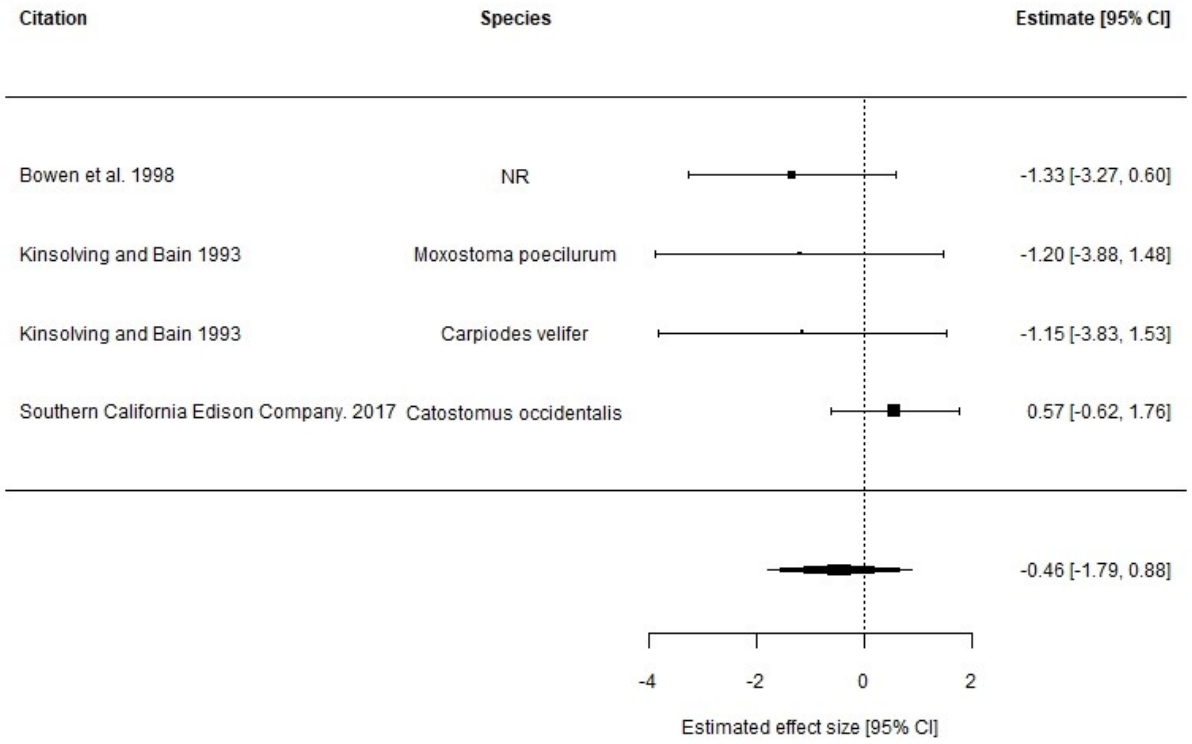

Fig. S130. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on the abundance of the family Catostomidae (*k*=4). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). NR: species not reported. Diamond: overall mean effect size of random-effects model.

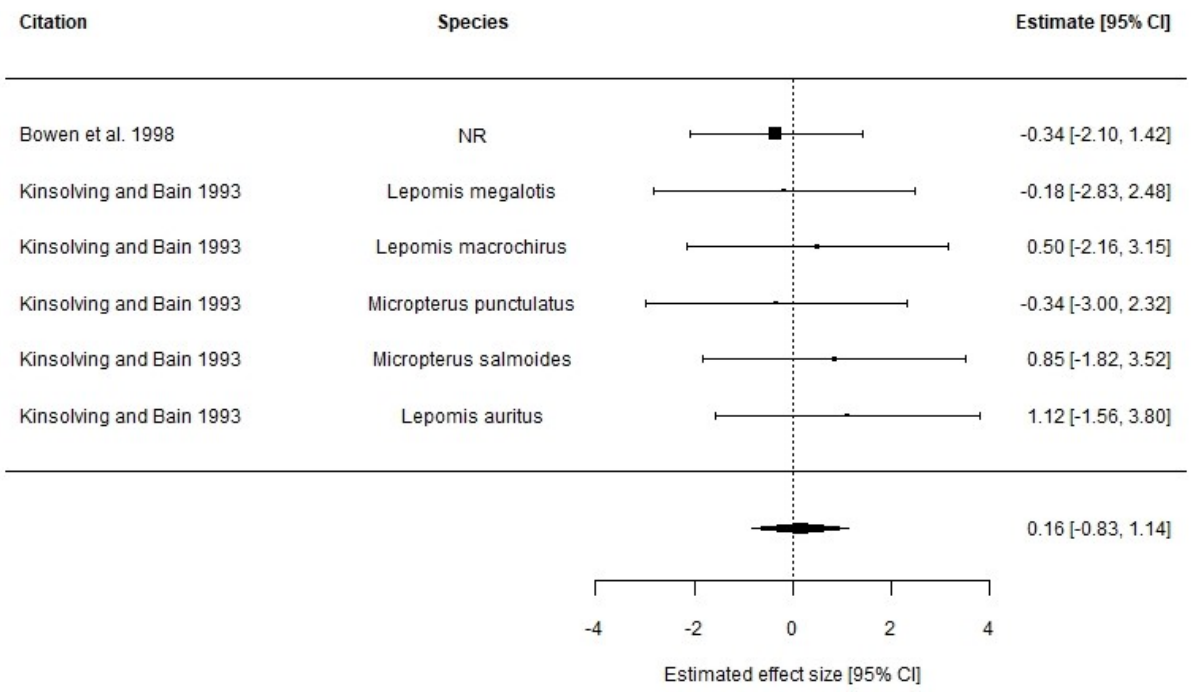

Fig. S31. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on the abundance of the family Centrarchidae (*k*=6). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). NR: species not reported. Diamond: overall mean effect size of random-effects model.

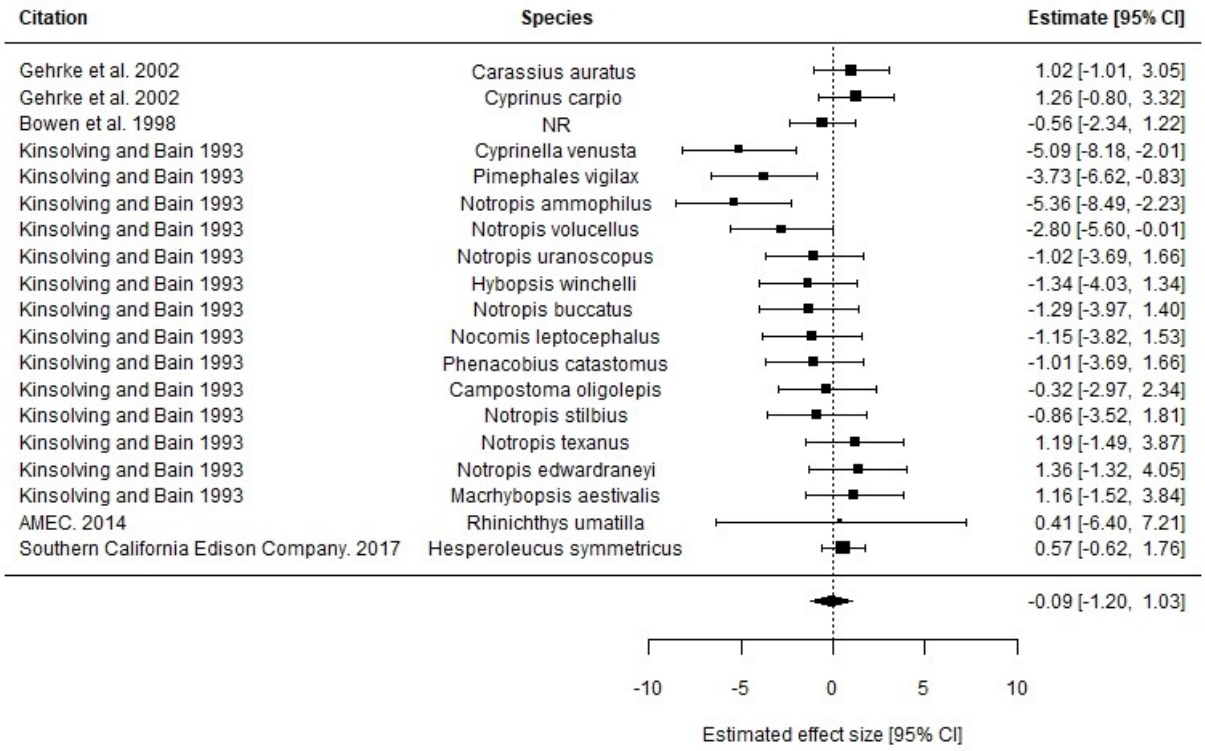

Fig. S32. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on the abundance of the family Cyprinidae (*k*=19). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). NR: species not reported. Diamond: overall mean effect size of random-effects model.

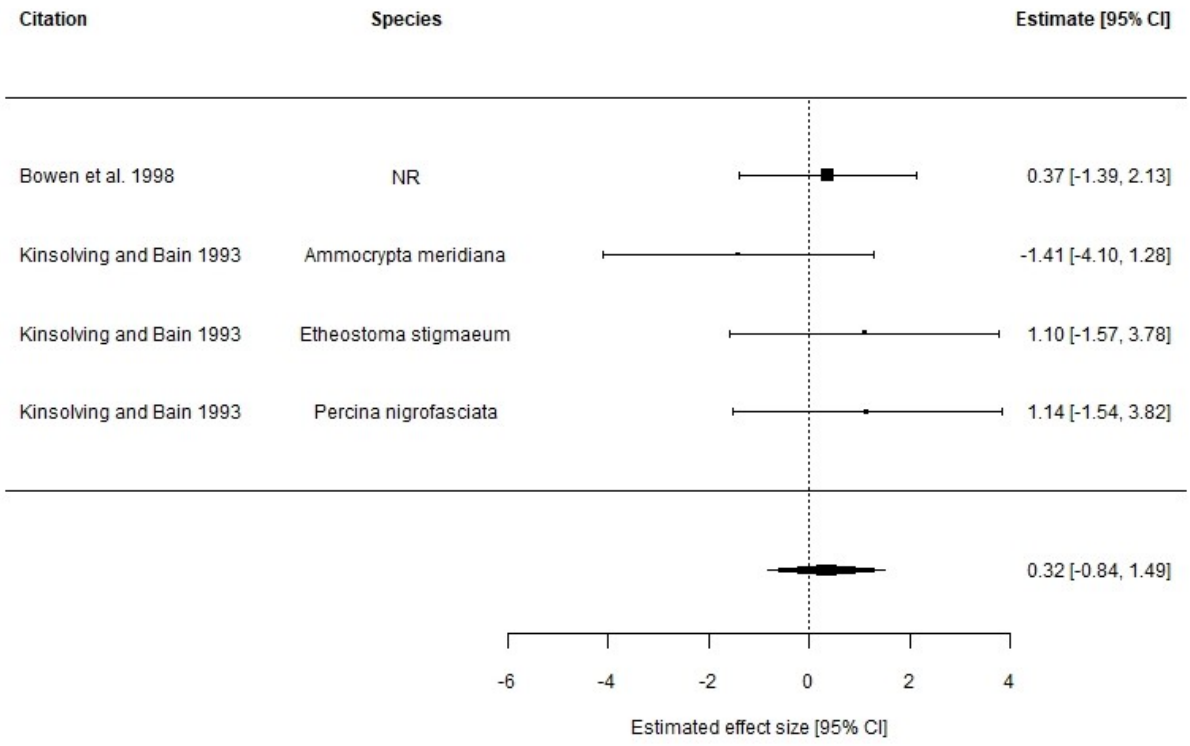

Fig. S33. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on the abundance of the family Percidae (*k*=4). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). NR: species not reported. Diamond: overall mean effect size of random-effects model.

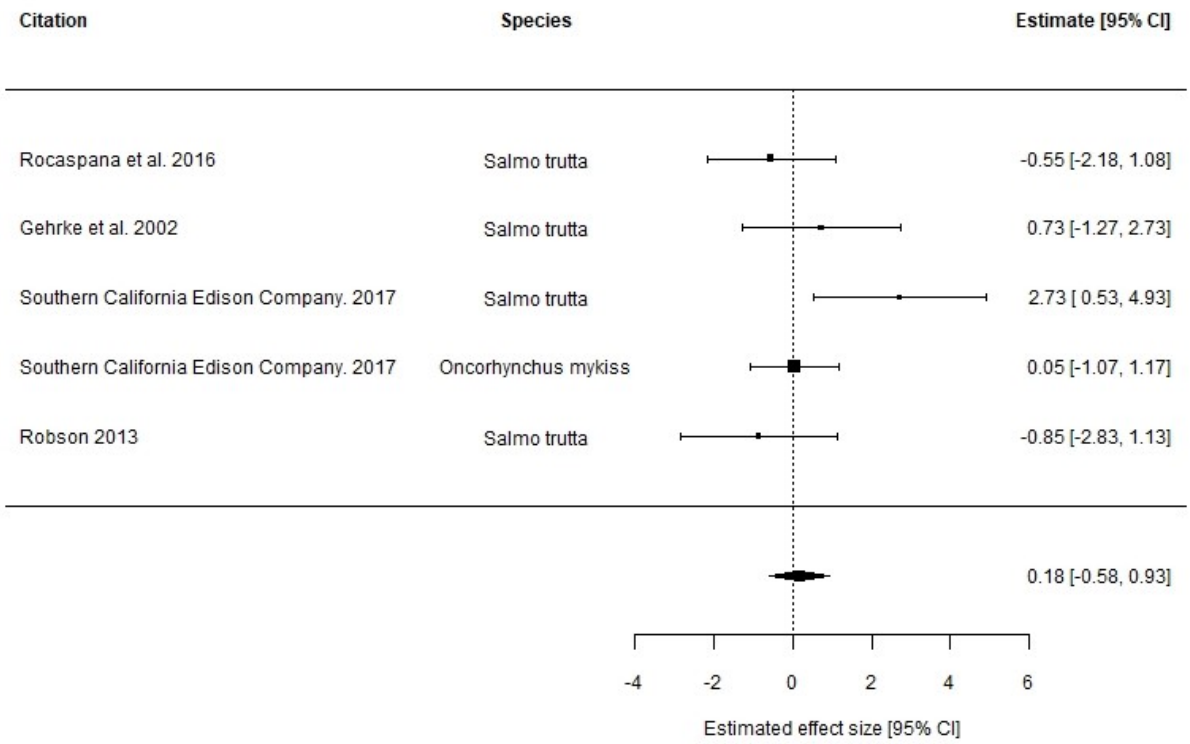

Fig. S34. Summary plot of all effect size estimates from *Control/Impact* evaluations of the impact of flow magnitude alterations on the abundance of the family Salmonidae (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in treatment areas than in comparator areas (no intervention). Diamond: overall mean effect size of random-effects model.

Within-year *Before/After* Studies: Family *Abundance*

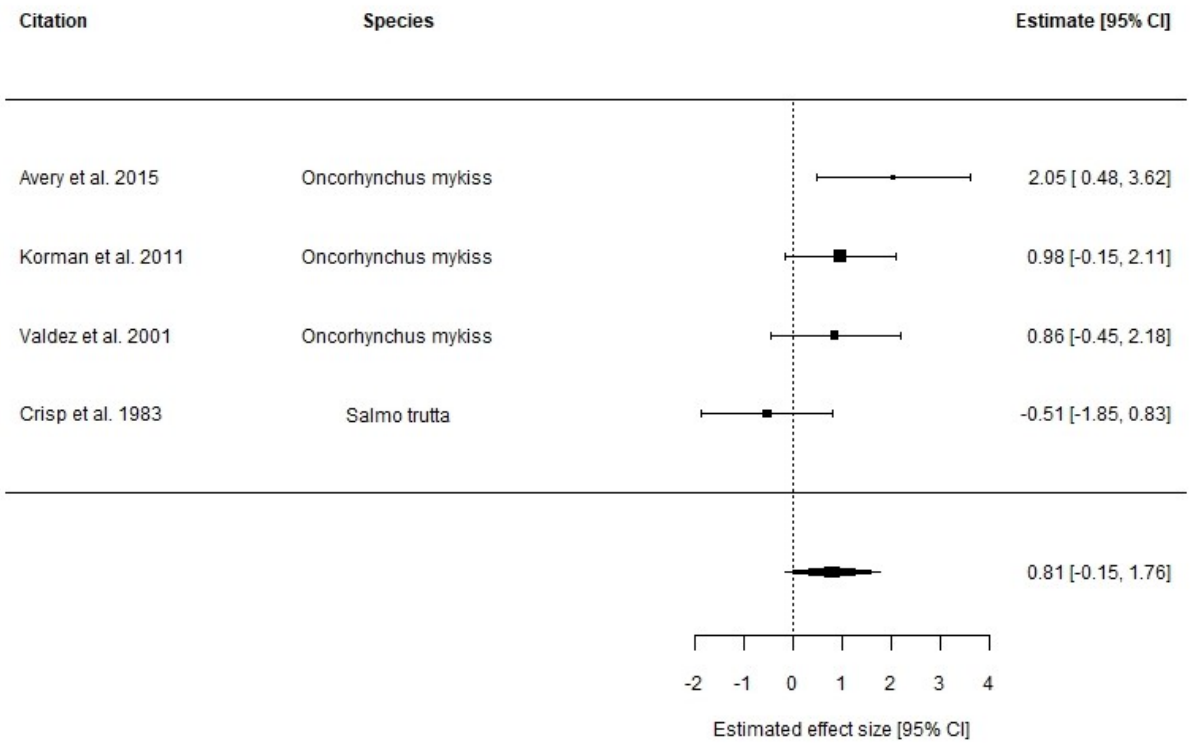

Fig. S35. Summary plot of all effect size estimates from with-in year *Before/After* studies considering post-intervention year-1 evaluations of the impact of flow magnitude alterations on the abundance of the family Salmonidae (*k*=4). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Interannual *Before/After* Studies: Family *Abundance*

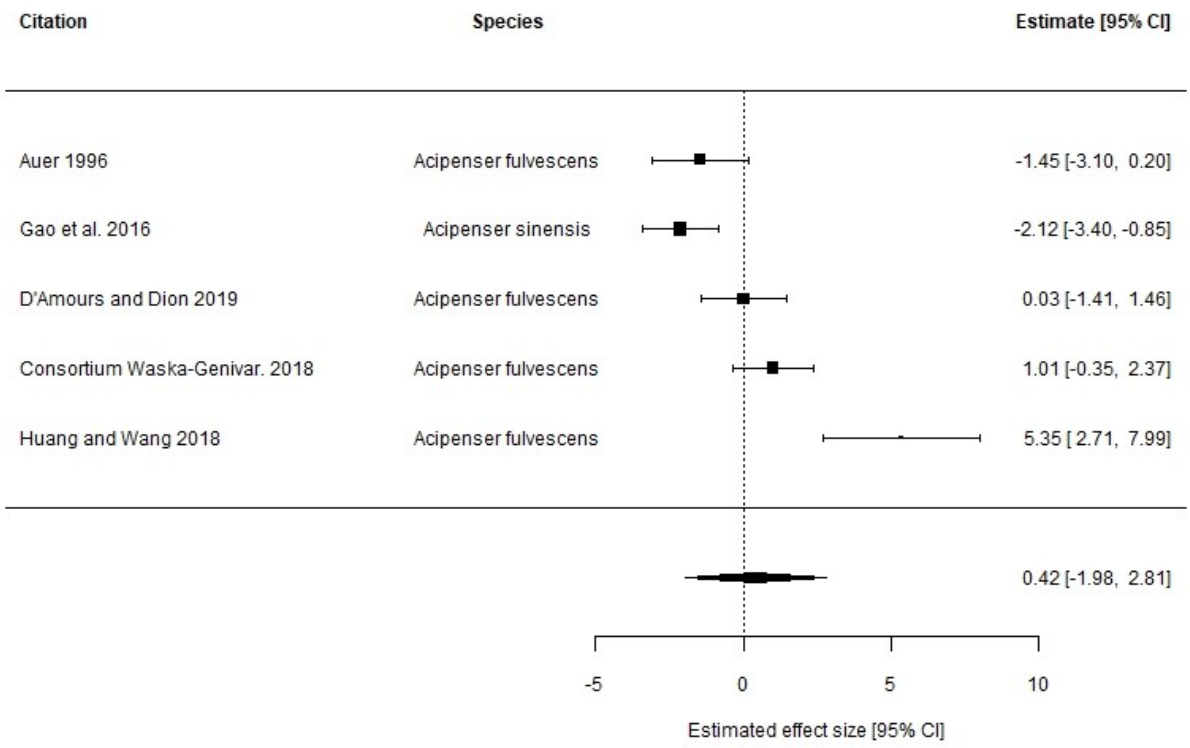

Fig. S36. Summary plot of all effect size estimates from interannual *Before/After* studies considering evaluations of the impact of flow magnitude alterations on the abundance of the family Acipenseridae (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

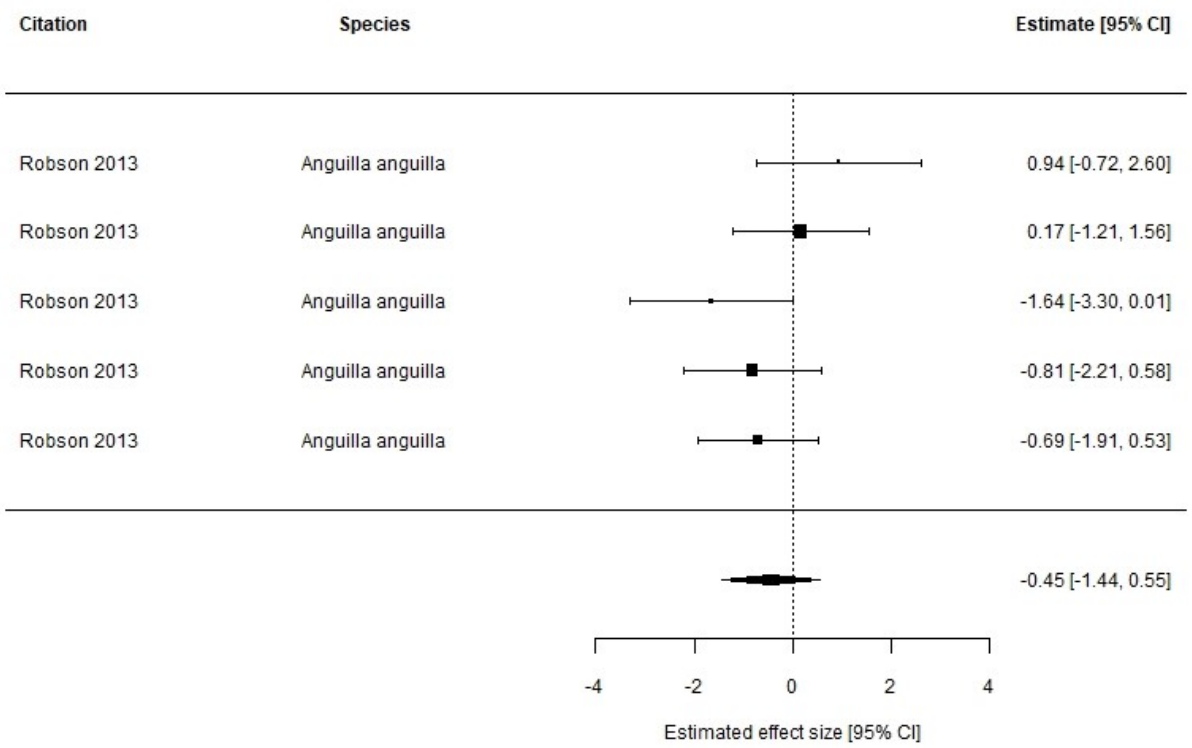

Fig. S37. Summary plot of all effect size estimates from interannual *Before/After* studies considering evaluations of the impact of flow magnitude alterations on the abundance of the family Anguillidae (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

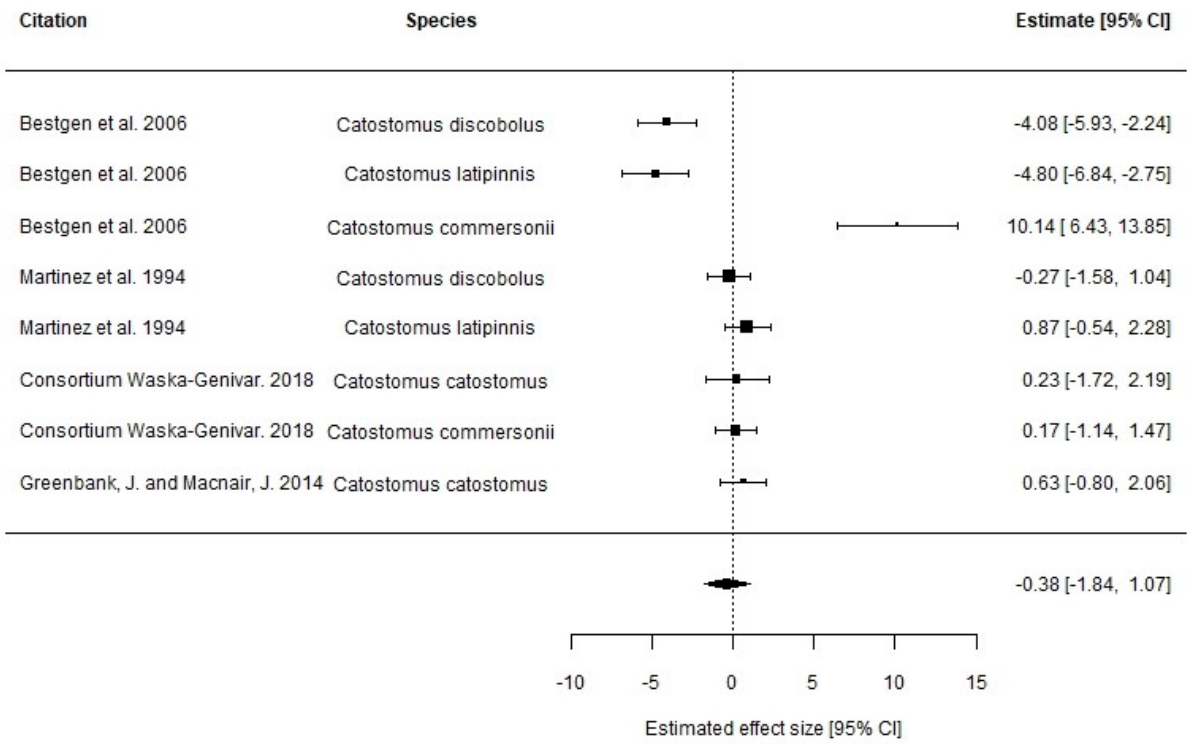

Fig. S38. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Catostomidae (*k*=8). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

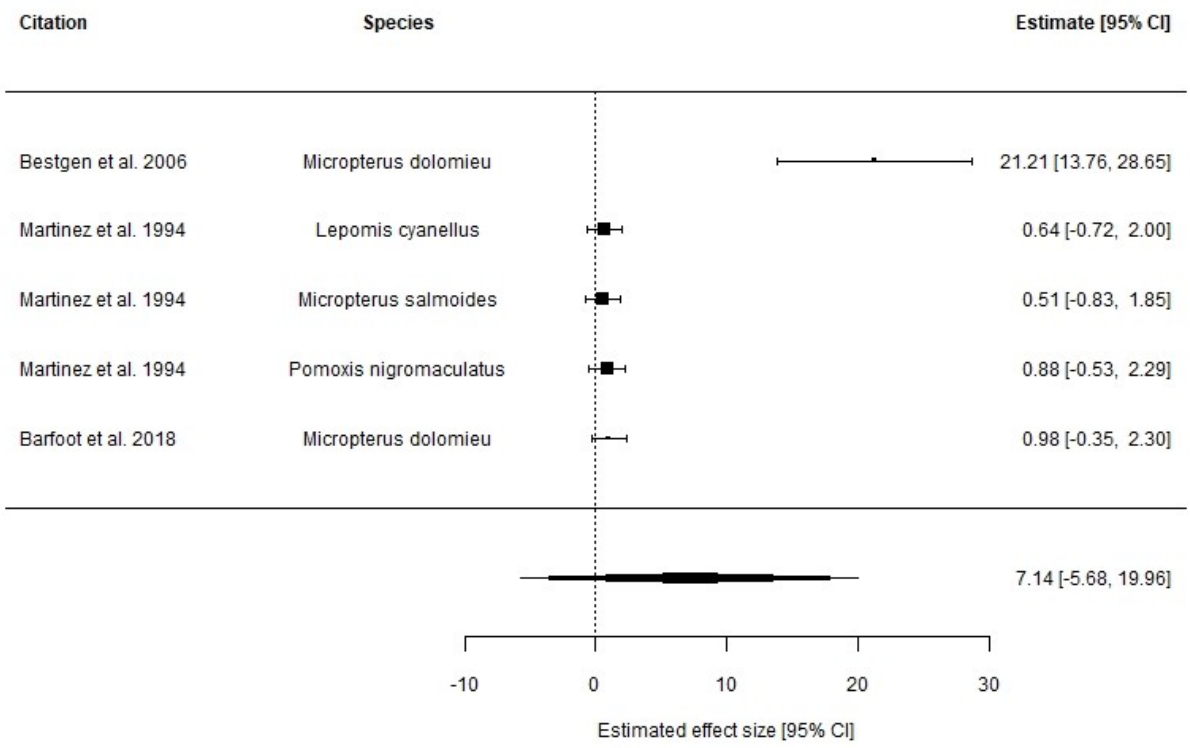

Fig. S39. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Centrarchidae (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

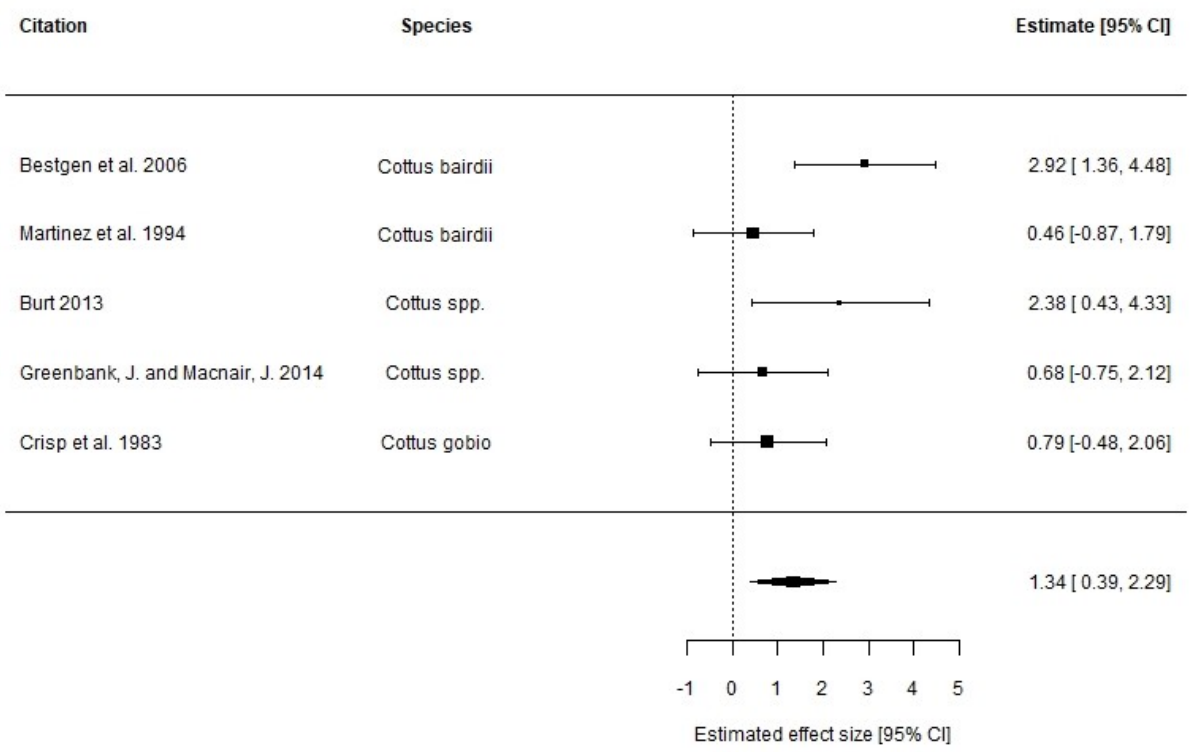

Fig. S40. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Cottidae (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

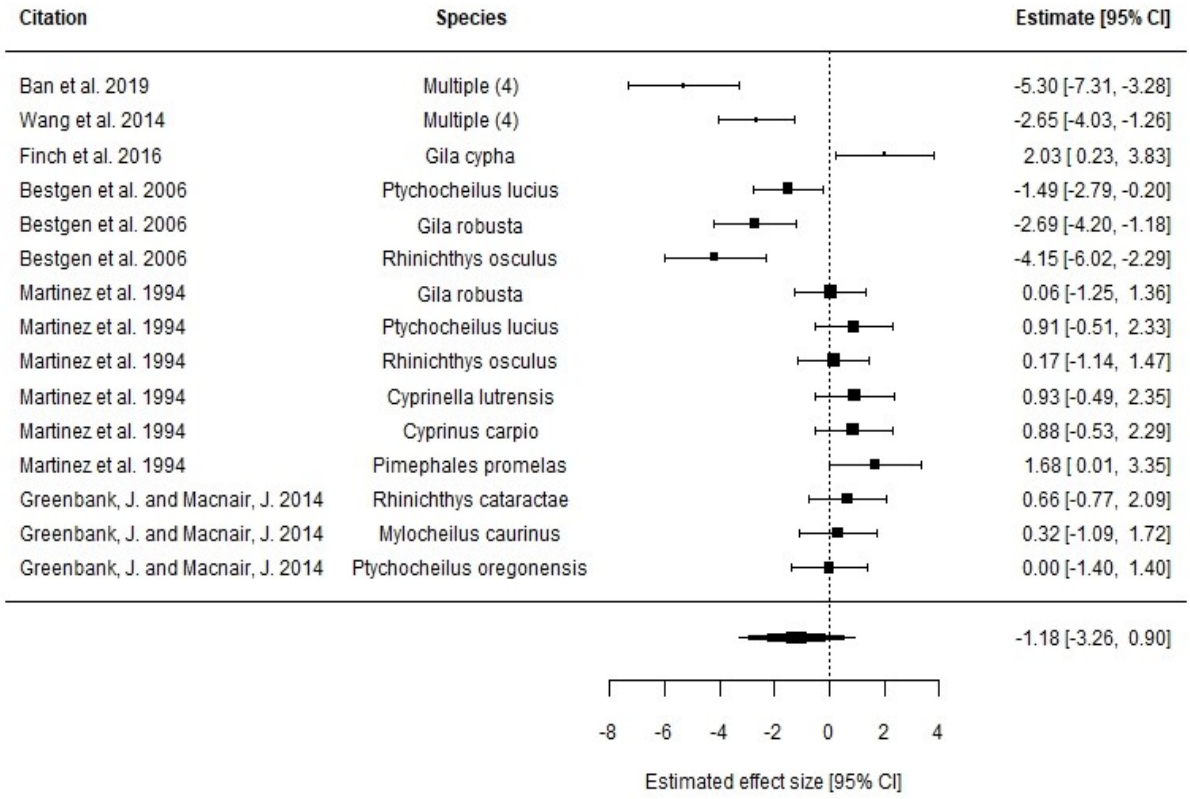

Fig. S41. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Cyprinidae (*k*=5). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model. *Multiple*: more than one species included in pooled data (number in brackets indicates number of species included).

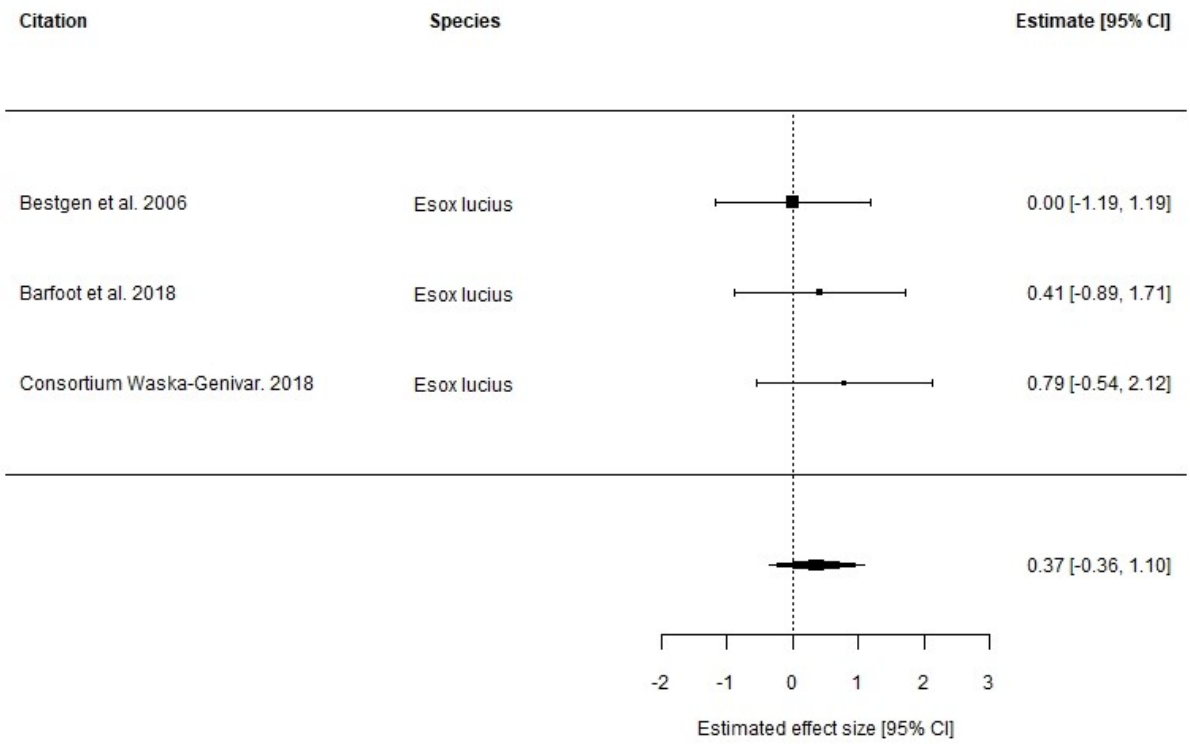

Fig. S42. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Esocidae (*k*=3). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

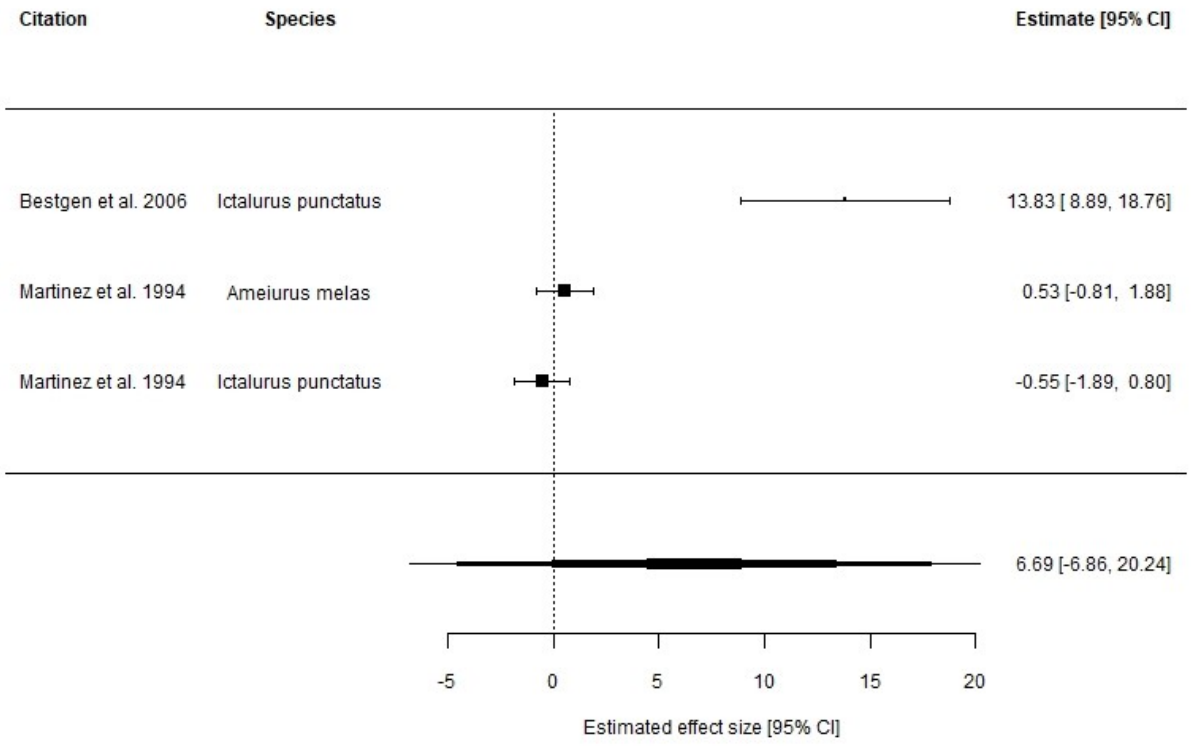

Fig. S43. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Ictaluridae (*k*=3). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

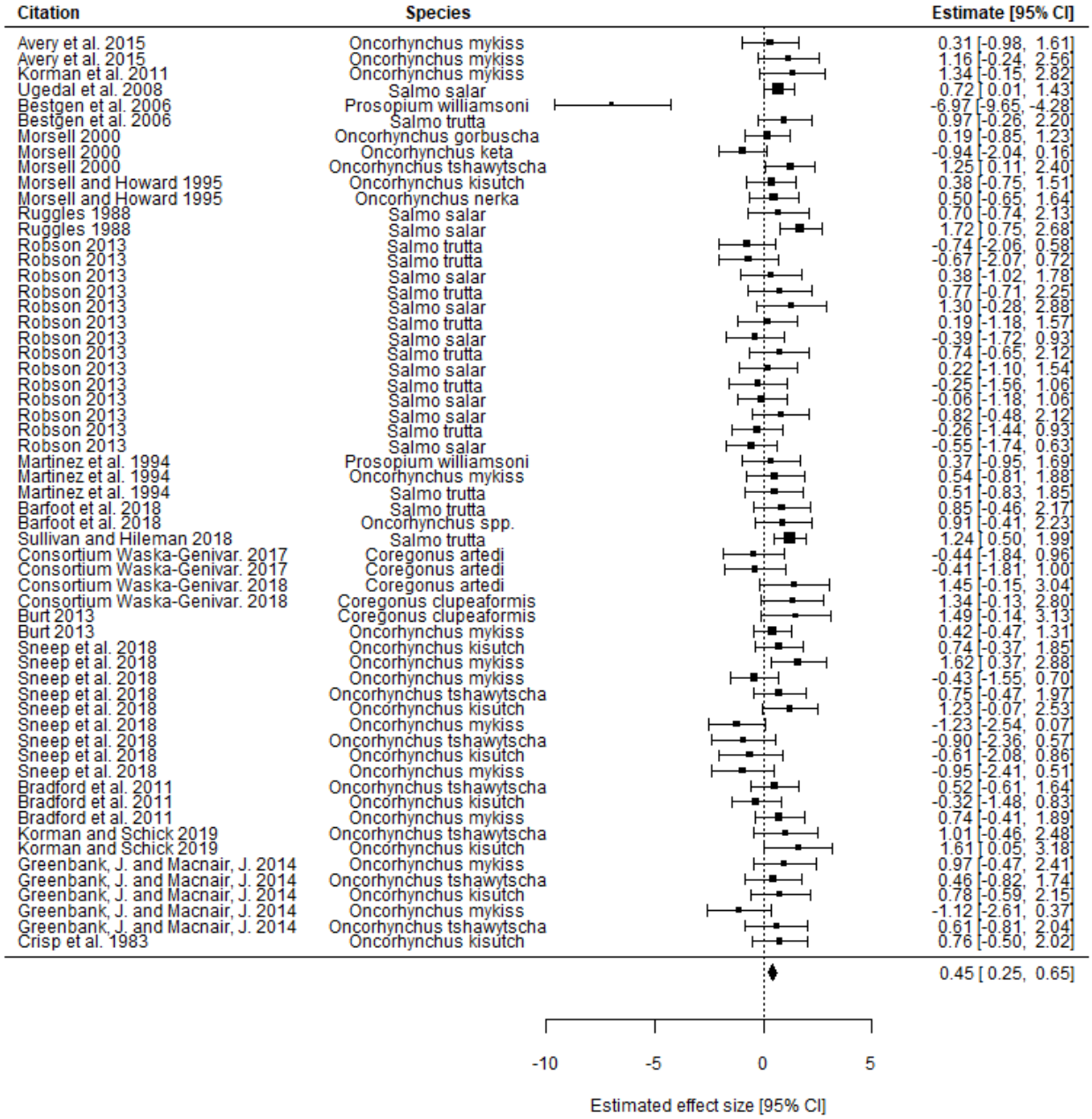

Fig. S44. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the family Salmonidae (*k*=59). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Biomass

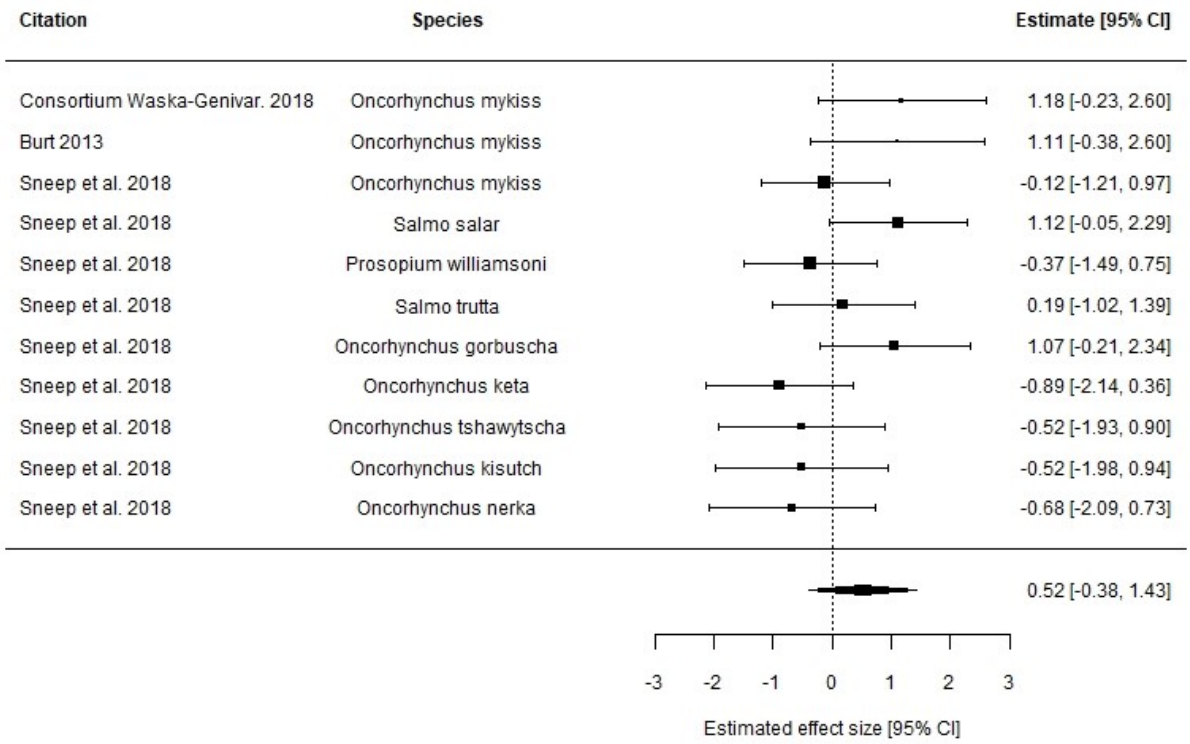

Fig. S45. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on biomass of the family Salmonidae (*k*=59). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Interannual *Before/After Studies*: Genera

Abundance

For families with statistically significant heterogeneity, we analyzed the response of genera therein, with sufficient sample size (i.e., ≥3 datasets from ≥2 independent studies) and variability to investigate responses in fish abundance to alterations in flow magnitude. Methods follow those used for other taxonomic analyses (see section in main text "Data synthesis and presentation – Quantitative synthesis").

There were only sufficient sample sizes to investigate genera within Cyprinidae and Salmonidae. Within Cyprinidae, there was sufficient sample size to investigate variation among three genera: (i) *Gila*; (ii) *Ptychocheilus*; and (iii) *Rhinichthys* (Fig S17-S19).

<u>*Gila:*</u>

- Average Hedge's g = -0.2311 (95% CI -2.8760, 2.4138; k = 3, p = 0.8640)
- (Q = 16.3347, p=0.0003)

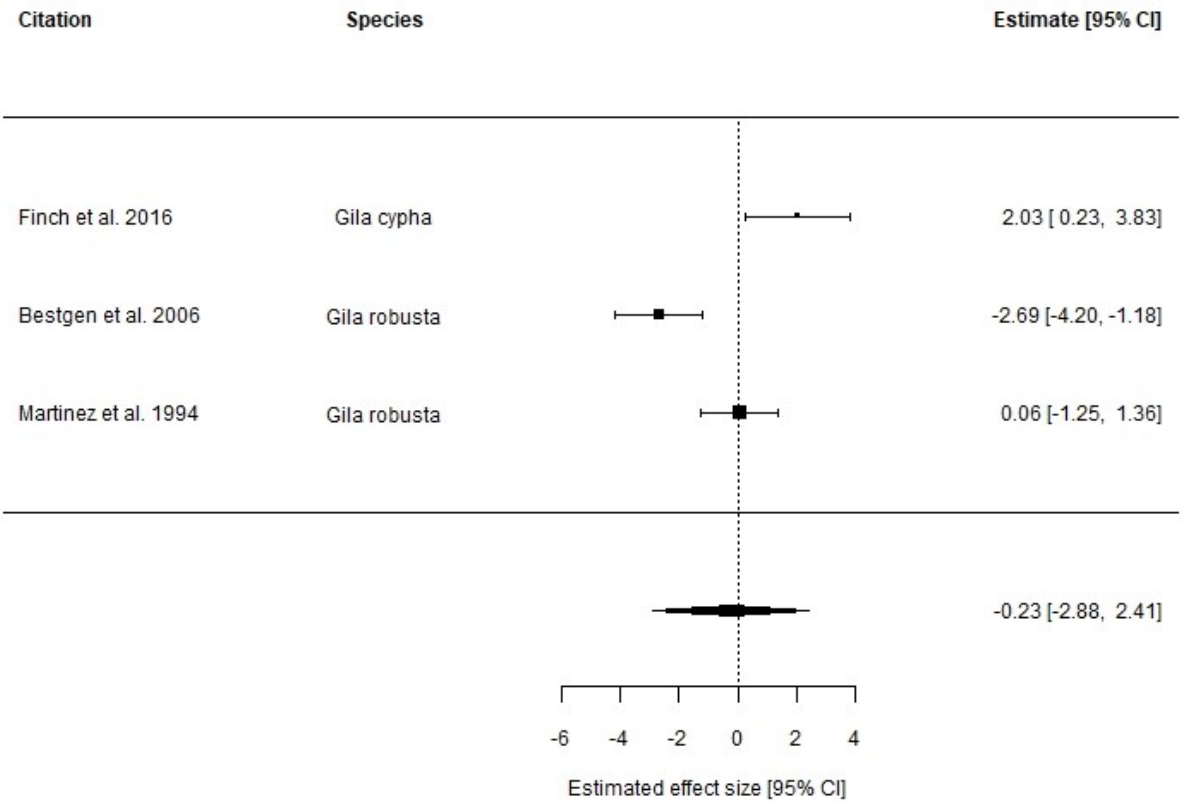

Fig. S46. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Gila* (k= 3). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Ptychocheilus:

- Average Hedge's g = -0.2197 (95% CI -1.6077, 1.1682; k = 3, p = 0.7564)
- (Q = 6.2363, p=0.0442)

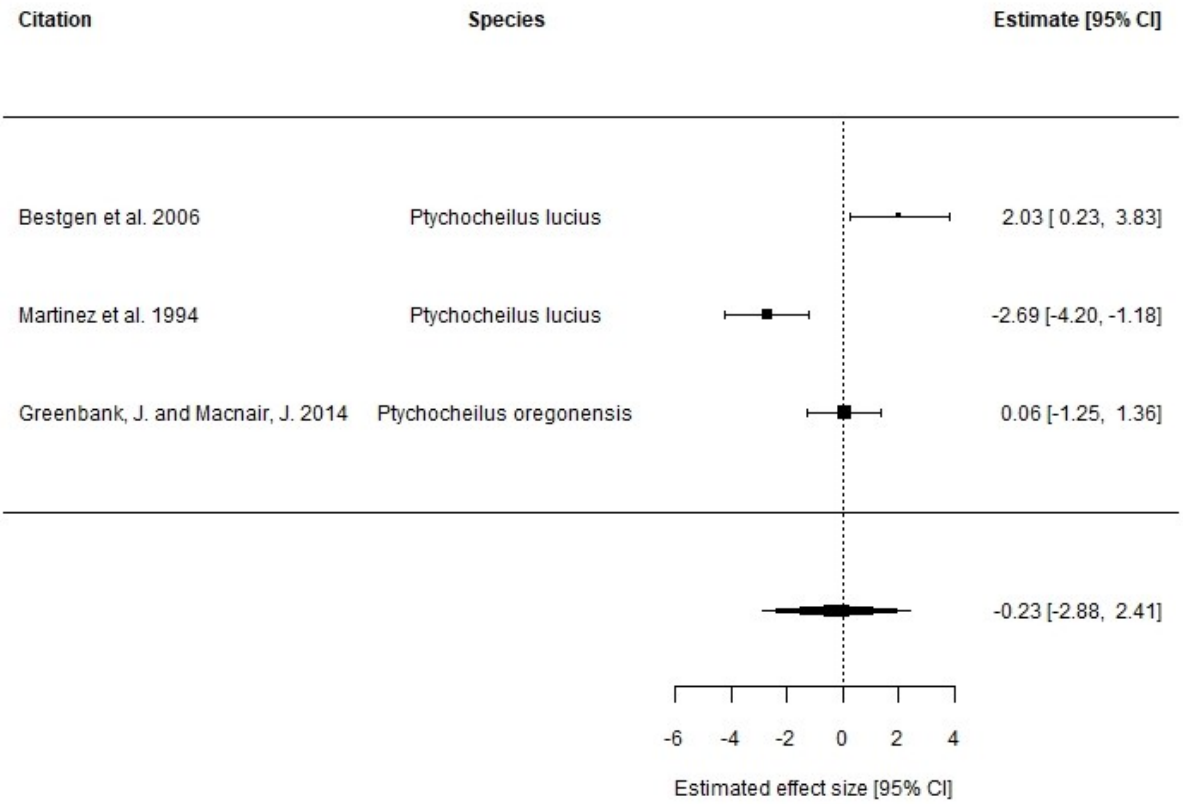

Fig. S47. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Ptychocheilus* (k= 3). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Rhinichthys

- Average Hedge's g = -1.0482 (95% CI -3.9785, 1.8821; k = 3, p = 0.4833)
- (Q = 18.2227, p=0.0001)

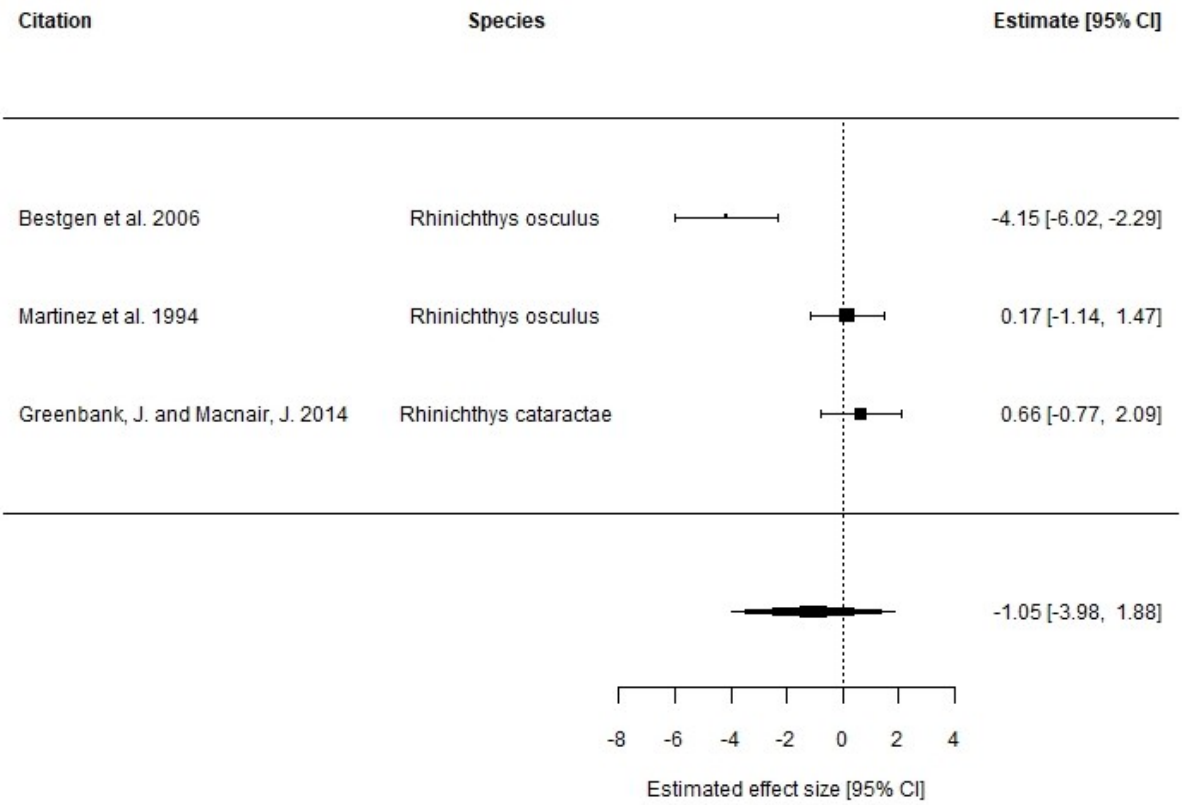

Fig. S48. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Rhinichthys* (k= 3). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Within Salmonidae, there was sufficient samples size to investigate variation among four genera: (i) *Coregonus*; (ii) *Oncorhynchus*; and (iii) *Prosopium*; and (iv) *Salmo* (Fig S20-S23).

<u>*Coregonus*</u>

- Average Hedge's g = 0.2526 (95% CI -0.9963, 1.5014; k = 4, p = 0.2526)
- (Q = 5.8771, p=0.1177)

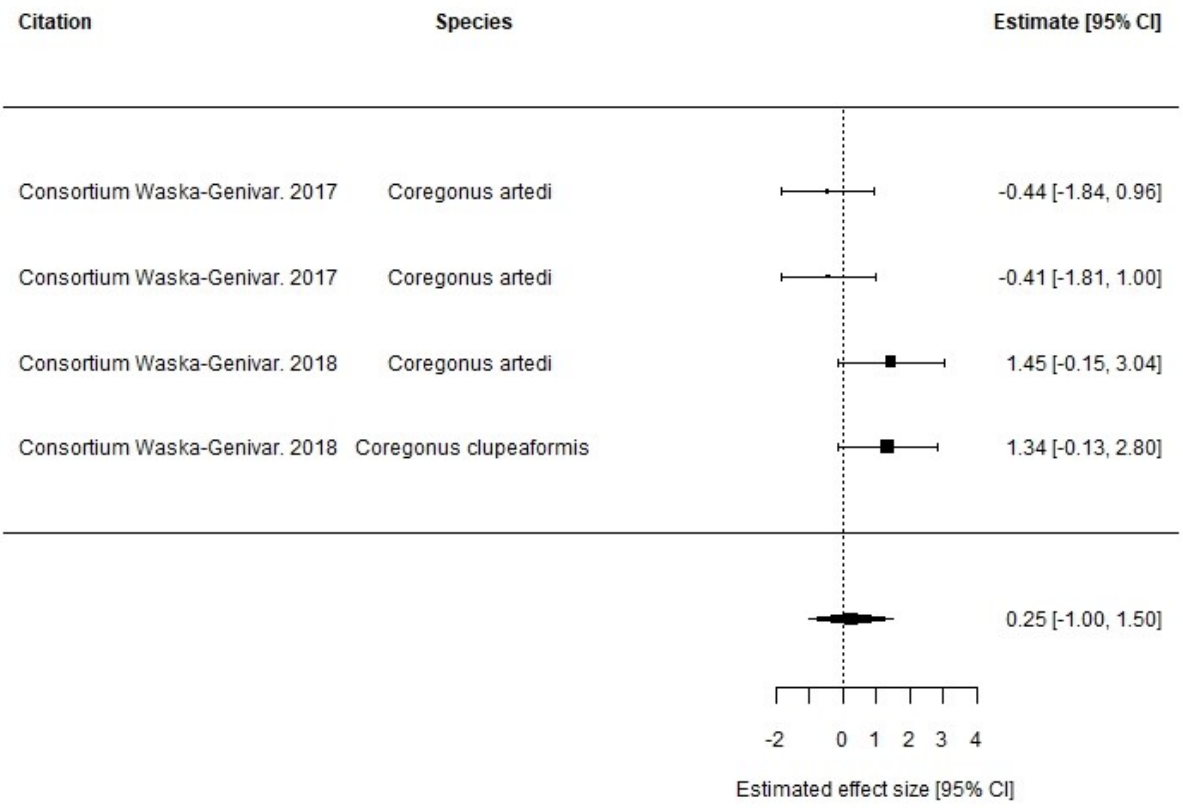

Fig. S49. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Coregonus* (k= 4). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Oncorhynchus

- Average Hedge's g = 0.3649 (95% CI 0.1335, 0.5962; k = 29, p = 0.0020)
- (Q = 42.4169, p=0.0396)

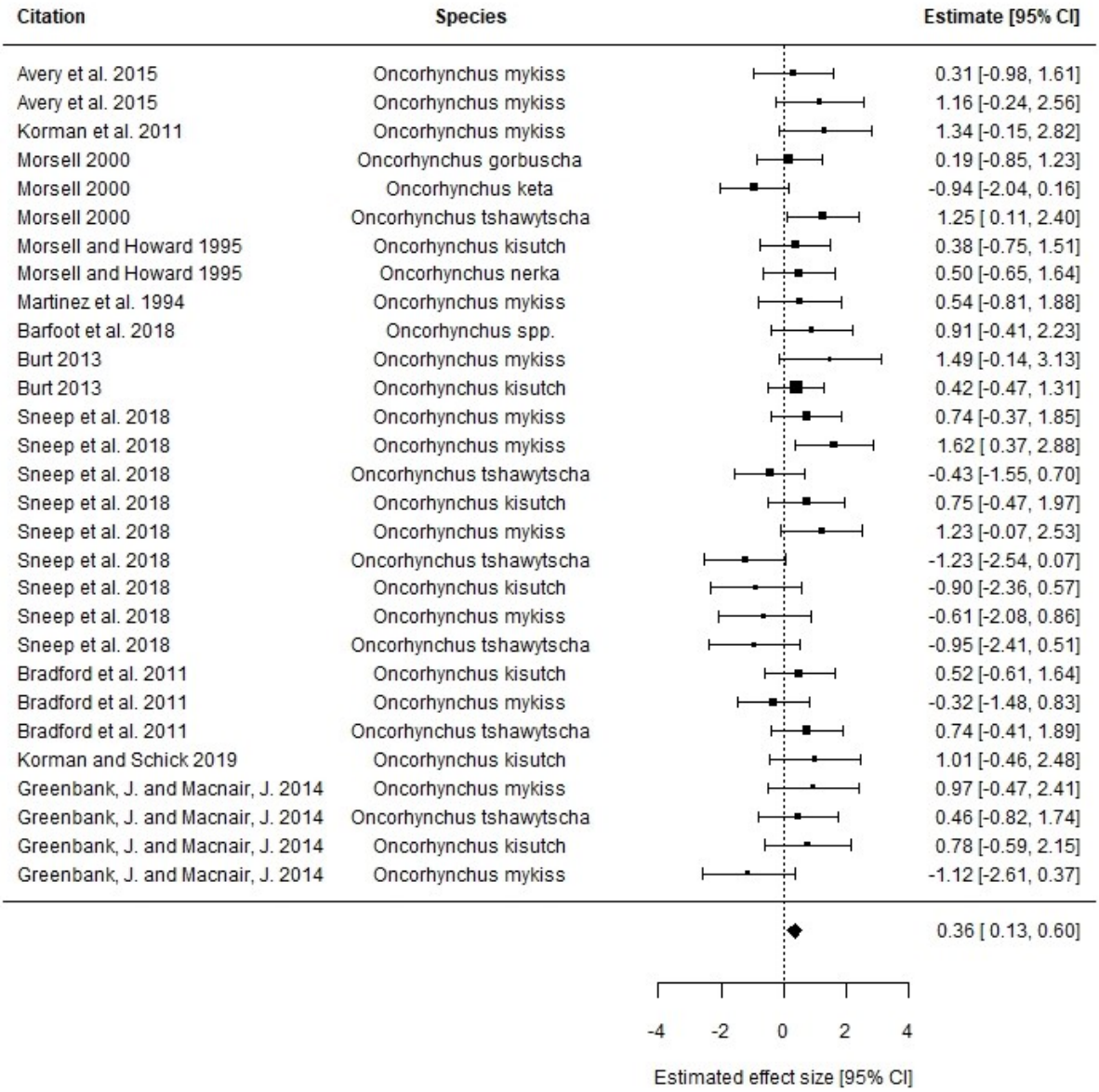

Fig. S50. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Oncorhynchus* (k= 29). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Salmo

- Average Hedge's g = 0.5305 (95% CI 0.1777, 0.8834; k = 22, p = 0.0032)
- (Q = 27.8024, p=0.1458)

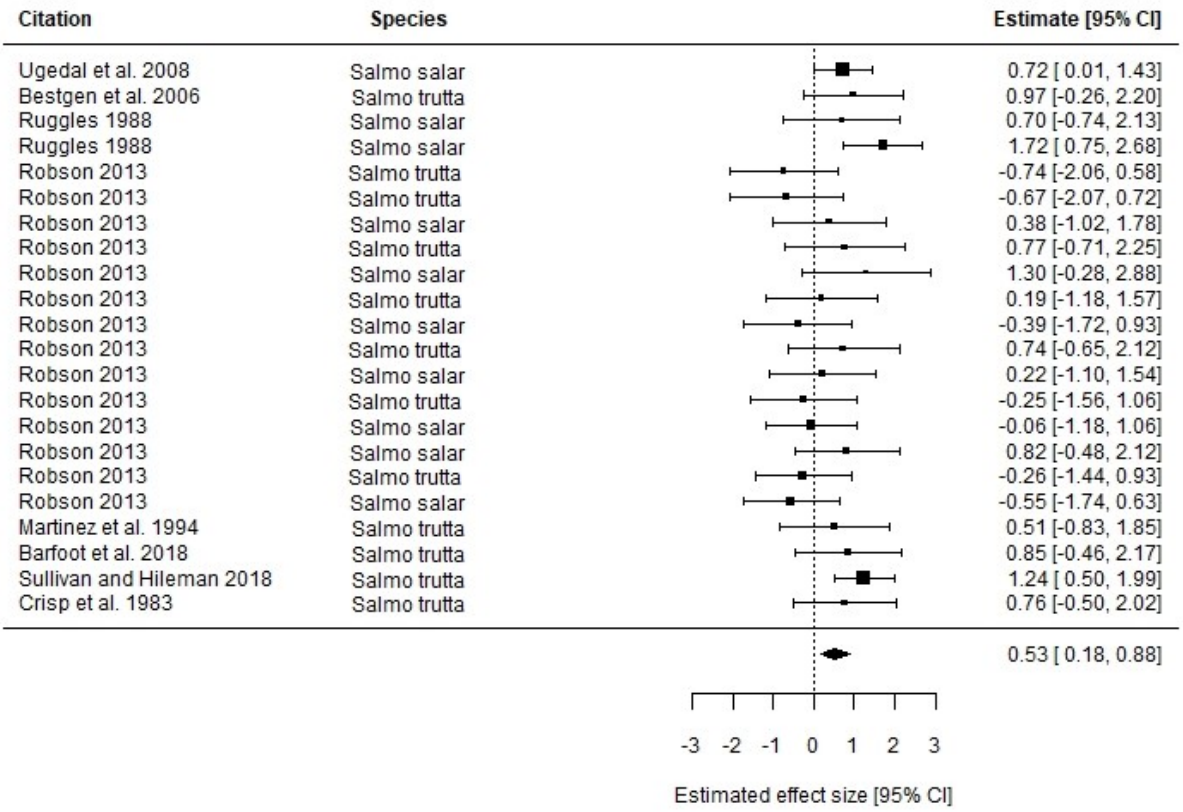

Fig. S51. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Salmo* (k= 22). Error bars indicate 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

Interannual *Before/After* Studies: Genera

Biomass

For families with statistically significant heterogeneity, we analyzed the response of genera therein, with sufficient sample size (i.e., ≥3 datasets from ≥2 independent studies) and variability to investigate responses in fish abundance to alterations in flow magnitude. Methods follow those used for other taxonomic analyses (see section in main text "Data synthesis and presentation – Quantitative synthesis").

There were only sufficient sample sizes to investigate genera within Salmonidae for the genus *Oncorhynchus.*

<u>*Oncorhynhcus*</u>

- Average Hedge's g = 0.3068 (95% CI -0.7191, 1.3326; k = 10, p = 0.5578)
- (Q = 12.7253, p=0.1754)

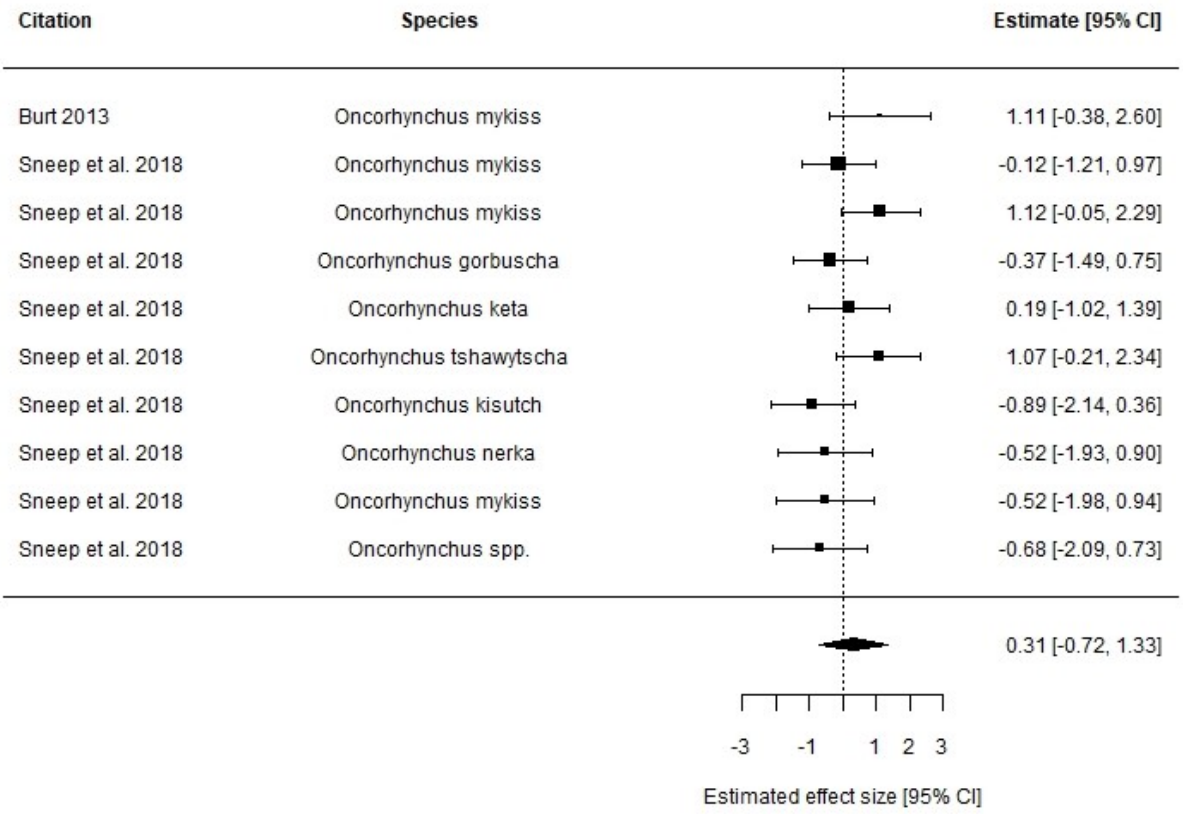

Fig. S52. Summary plot of all effect size estimates from interannual *Before/After* studies considering the impact of flow magnitude alterations on abundance of the genus *Oncorhynchus* (k= 10). Error bars indicated 95% confidence intervals. A positive mean value (right of dashed zero line) indicates that the abundance was higher in *After* period than in *Before* period (no intervention). Diamond: overall mean effect size of random-effects model.

References

Acreman, M., A. H. Arthington, M. J. Colloff, C. Couch, N. D. Crossman, F. Dyer, I. Overton, C. A. Pollino, M. J. Stewardson, and W. Young. 2014. Environmental flows for natural, hybrid, and novel riverine ecosystems in a changing world. Frontiers in Ecology and the Environment 12(8):466–473.

Algera, D. A., T. Rytwinski, J. J. Taylor, J. R. Bennett, K. E. Smokorowski, P. M. Harrison, K. D. Clarke, E. C. Enders, M. Power, M. S. Bevelhimer, and S. J. Cooke. 2020. What are the relative risks of mortality and injury for fish during downstream passage at hydroelectric dams in temperate regions? A systematic review. Environmental Evidence 9(1):3.

Arthington, A. H., S. E. Bunn, N. L. Poff, and R. J. Naiman. 2006. The challenge of providing environmental flow rules to sustain river ecosystems. Ecological Applications 16(4):1311–1318.

Avery, L. A., J. Korman, and W. R. Persons. 2015. Effects of increased discharge on spawning and age-0 recruitment of rainbow trout in the Colorado River at Lees Ferry, Arizona. North American Journal of Fisheries Management 35(4):671–680.

Barrett, T. J. B. J., and K. R. M. R. Munkittrick. 2010. Seasonal reproductive patterns and recommended sampling times for sentinel fish species used in environmental effects monitoring programs in Canada. Environmental Reviews 18:115-135.

Bernes, C., B. Macura, B. G. Jonsson, K. Junninen, J. Müller, J. Sandström, A. Lõhmus, and E. Macdonald. 2018. Manipulating ungulate herbivory in temperate and boreal forests: effects on vegetation and invertebrates. A systematic review. Environmental Evidence 7(1):13.

Bestgen, K. R., K. A. Zelasko, R. I. Compton, and T. Chart. 2006. Responses of the Green River fish community to changes in flow and temperature regimes from Flaming Gorge Dam since 1996 based on sampling conducted from 2002 to 2004. Larval Fish Laboratory, Colorado State University, Final Report to the Colorado River Recovery Implementation Program, Project 115, Denver.

Bilotta, G. S., A. M. Milner, and I. Boyd. 2014. On the use of systematic reviews to inform environmental policies. Environmental Science & Policy 42:67–77.

Birnie-Gauvin, K., T. Rytwinski, M. Harper, J. Taylor, A. Smith, K. E. Smokorowski, K. Turgeon, and S. J. Cooke. In review. How do natural changes in flow magnitude affect fish abundance and diversity in temperate regions? A systematic review protocol. Ecological Solutions and Evidence 0:000-000.

Block, B. D., B. A. Denfeld, J. D. Stockwell, G. Flaim, H.-P. F. Grossart, L. B. Knoll, D. B. Maier, R. L. North, M. Rautio, J. A. Rusak, S. Sadro, G. A. Weyhenmeyer, A. J. Bramburger, D. K. Branstrator, K. Salonen, and S. E. Hampton. 2019. The unique methodological challenges of winter limnology. Limnology and Oceanography: Methods 17(1):42–57.

Block, B. D., J. D. Stockwell, and J. E. Marsden. 2020. Contributions of winter foraging to the annual growth of thermally dissimilar fish species. Hydrobiologia 847(20):4325–4341.

Boavida, I., M. Díaz-Redondo, J. F. Fuentes-Pérez, D. S. Hayes, J. Jesus, M. Moreira, O. Belmar, N. Vila-Martínez, A. Palau-Nadal, and M. J. Costa. 2020. Ecohydraulics of river flow alterations and impacts on freshwater fish. Limnetica.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009a. Recreating the summary data for the full study (option 2). Pages 221–222 *in* M. Borenstein, L. V.

Hedges, J. P. T. Higgins, and H. R. Rothstein, editors. Introduction to meta-analysis. Johgn Wiley & Sons, Ltd., United Kingdom.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009b. Effect sizes based on means. Pages 21–32 *in* M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, editors. Introduction to meta-analysis. John Wiley & Sons, Ltd, United Kingdom.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009c. Multiple outcomes or time-points within a study. Pages 225–238 *in* M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, editors. Introduction to meta-analysis. John Wiley & Sons, Ltd, United Kingdom.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Research Synthesis Methods 1(2):97–111.

Borgstroem, R., and O. Skaala. 1993. Size-dependent catchability of brown trout and Atlantic salmon parr by electrofishing in a low conductivity stream. Nordic journal of freshwater research 6:14–20.

Bowen, Z. H., M. C. Freeman, and K. D. Bovee. 1998. Evaluation of Generalized Habitat Criteria for Assessing Impacts of Altered Flow Regimes on Warmwater Fishes. Transactions of the American Fisheries Society 127(3):455–468. Taylor & Francis.

Brady, M. E., A. M. Chione, and J. B. Armstrong. 2020. Missing pieces in the full annual cycle of fish ecology: a systematic review of the phenology of freshwater fish research. bioRxiv:2020.11.24.395665.

Bunn, S. E., and A. H. Arthington. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental Management 30(4):492–507.

Christie, A. P., T. Amano, P. A. Martin, G. E. Shackelford, B. I. Simmons, W. J. Sutherland. 2019. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. Journal of Applied Ecology. 56 (12):2742-2754.

Cohen, J. 1977. Statistical power analysis for the behavioural sciences. Revised edition. Academic Press, New York, New York.

[CEE] Collaboration for Environmental Evidence. 2018. Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.0.

Collins, A. M., D. Coughlin, and N. Randall. 2019. Engaging environmental policy-makers with systematic reviews: challenges, solutions and lessons learned. Environmental Evidence 8(1):2.

Cook, C. N., H. P. Possingham, and R. A. Fuller. 2013. Contribution of Systematic Reviews to Management Decisions. Conservation Biology 27(5):902–915.

Cooke, S. J., J. C. Rice, K. A. Prior, R. Bloom, O. Jensen, D. R. Browne, L. A. Donaldson, J. R. Bennett, J. C. Vermaire, and G. Auld. 2016. The Canadian context for evidence-based conservation and environmental management. Environmental Evidence 5(1):14.

Counihan, T. D., I. R. Waite, A. F. Casper, D. L. Ward, J. S. Sauer, E. R. Irwin, C. G. Chapman, B. S. Ickes, C. P. Paukert, J. J. Kosovich, and J. M. Bayer. 2018. Can data from disparate long-term fish monitoring programs be used to increase our understanding of regional and continental trends in large river assemblages? PLOS ONE 13(1):e0191472.

Coutant, C. C., and R. R. Whitney. 2000. Fish Behavior in Relation to Passage through Hydropower Turbines: A Review. Transactions of the American Fisheries Society 129(2):351–380.

Couto, T. B., and J. D. Olden. 2018. Global proliferation of small hydropower plants – science and policy. Frontiers in Ecology and the Environment 16(2):91–100.

Crisp, D. T., R. H. K. Mann, and P. R. Cubby. 1983. Effects of regulation of the River Tees upon fish populations below Cow Green Reservoir. Journal of Applied Ecology 20(2):371–386.

Drahota, A., and E. Bellor. 2008. RevMan Calculator: Finding standard deviations. The Cochrane Collaboration.

Eales, J., N. R. Haddaway, C. Bernes, S. J. Cooke, B. G. Jonsson, J. Kouki, G. Petrokofsky, and J. J. Taylor. 2018. What is the effect of prescribed burning in temperate and boreal forest on biodiversity, beyond pyrophilous and saproxylic species? A systematic review. Environmental Evidence 7(1):19.

Enders, E. C., D. A. Watkinson, H. Ghamry, K. H. Mills, and W. G. Franzin. 2017. Fish age and size distributions and species composition in a large, hydropeaking Prairie River. River Research and Applications 33(8):1246–1256.

Fricke, R., W. Eschmeyer, and R. Van der Laan, editors. 2020. Eschemeyer's catalog of fishes: genera, species, references (http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp).

Froese, R., and D. Pauly, editors. 2020. FishBase. World Wide Web electronic publication. www.fishbase.org, version (12/2020).

Fuller, M. R., M. W. Doyle, and D. L. Strayer. 2015. Causes and consequences of habitat fragmentation in river networks. Annals of the New York Academy of Sciences 1355(1):31–51.

Gillespie, B. R., S. Desmet, P. Kay, M. R. Tillotson, and L. E. Brown. 2015. A critical analysis of regulated river ecosystem responses to managed environmental flows from reservoirs. Freshwater Biology 60(2):410–425.

Göthe, E., E. Degerman, L. Sandin, J. Segersten, C. Tamario, and B. G. Mckie. 2019. Flow restoration and the impacts of multiple stressors on fish communities in regulated rivers. Journal of Applied Ecology 56(7):1687–1702.

Grill, G., B. Lehner, M. Thieme, B. Geenen, D. Tickner, F. Antonelli, S. Babu, P. Borrelli, L. Cheng, H. Crochetiere, H. E. Macedo, R. Filgueiras, M. Goichot, J. Higgins, Z. Hogan, B. Lip, M. E. McClain, J. Meng, M. Mulligan, C. Nilsson, J. D. Olden, J. J. Opperman, P. Petry, C. R. Liermann, L. Sáenz, S. Salinas-Rodríguez, P. Schelle, R. J. P. Schmitt, J. Snider, F. Tan, K. Tockner, P. H. Valdujo, A. van Soesbergen, and C. Zarfl. 2019. Mapping the world's free-flowing rivers. Nature 569(7755):215–221.

Guénard, G., G. Lanthier, S. Harvey-Lavoie, C. J. Macnaughton, C. Senay, M. Lapointe, P. Legendre, and D. Boisclair. 2016. A spatially-explicit assessment of the fish population response to flow management in a heterogeneous landscape. Ecosphere 7(5):e01252.

Gusenbauer, M., and N. R., Haddaway. 2019. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research Synthesis Methods 11 (2): 181-217.

Haddaway, N. R., B. Macura, P. Whaley, and A. S. Pullin. 2017. ROSES for systematic review reports. Version 1.0.

Haddaway, N. R., B. Macura, P. Whaley, and A. S. Pullin. 2018. ROSES RepOrting Standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. Environmental Evidence 7(1):7.

Hampton, S. E., M. D. Scheuerell, M. J. Church, and J. M. Melack. 2019. Long-term perspectives in aquatic research. Limnology and Oceanography 64(S1):S2–S10.

Harper, M., T. Rytwinski, J. J. Taylor, J. R. Bennett, K. E. Smokorowski, and S. J. Cooke. 2020. How do changes in flow magnitude due to hydroelectric power production affect fish abundance and diversity in temperate regions? A systematic review protocol. Environmental Evidence 9(1):14.

Hedger, R. D., O. H. Diserud, O. T. Sandlund, L. Saksgård, O. Ugedal, and G. Bremset. 2018. Bias in estimates of electrofishing capture probability of juvenile Atlantic salmon. Fisheries Research 208:286–295.

Hedges, L. V. 1981. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics 6(2):107–128.

Hedges, L. V., E. Tipton, and M. C. Johnson. 2010. Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods 1(1):39–65.

Hedges, L., and J. Vevea. 1998. Fixed- and random-effects models in meta-analysis. Psychological Methods 3(4):486–504.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54(2):187–211.

[IHA] International Hydropower Association. 2020. 2020 Hydropower Status Report. International Hydropower Association, London.

Kadykalo, A. N., N. R. Haddaway, T. Rytwinski, and S. J. Cooke. 2021. Ten principles for generating accessible and useable COVID-19 environmental science and a fit-for-purpose evidence base. Ecological Solutions and Evidence 2(1):e12041.

Kambach, S., H. Bruelheide, K. Gerstner, J. Gurevitch, M. Beckmann, and R. Seppelt. 2020. Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. Ecology and Evolution 10(20):11699–11712.

Kinsolving, A. D., and M. B. Bain. 1993. Fish assemblage recovery along a riverine disturbance gradient. Ecological Applications: A Publication of the Ecological Society of America 3(3):531–544.

Konrad, C. P., J. D. Olden, D. A. Lytle, T. S. Melis, J. C. Schmidt, E. N. Bray, M. C. Freeman, K. B. Gido, N. P. Hemphill, M. J. Kennard, L. E. McMullen, M. C. Mims, M. Pyron, C. T. Robinson, and J. G. Williams. 2011. Large-scale flow experiments for managing river systems. BioScience 61(12):948–959.

Korman, J., C. Walters, S. J. D. Martell, W. E. Pine, and Dutterer. 2011. Effects of flow fluctuations on habitat use and survival of age-0 rainbow trout (*Oncorhynchus mykiss*) in a large, regulated river. Canadian Journal of Fisheries and Aquatic Sciences 68:1097-1109.

Light, R. J., and D. B. Pillemer. 1984. Summing up: The science of reviewing research. Harvard University Press, Cambridge, United States.

Lipsey, M. W. 2003. Those confounded moderators in meta-analysis: Good, bad, and ugly. The Annals of the American Academy of Political and Social Science 587:69–81.

Lohner, T. W., and D. A. Dixon. 2013. The value of long-term environmental monitoring programs: an Ohio River case study. Environmental Monitoring and Assessment 185(11):9385–9396.

Lytle, D. A., and N. L. Poff. 2004. Adaptation to natural flow regimes. Trends in Ecology & Evolution 19(2):94–100.

Macura, B., P. Byström, L. Airoldi, B. K. Eriksson, L. Rudstam, and J. G. Støttrup. 2019. Impact of structural habitat modifications in coastal temperate systems on fish recruitment: a systematic review. Environmental Evidence 8(1):14.

Martin, P. A., G. E. Shackelford, J. M. Bullock, B. Gallardo, D. C. Aldridge, and W. J. Sutherland. 2020. Management of UK priority invasive alien plants: a systematic review protocol. Environmental Evidence 9(1):1.

McManamay, R. A., D. J. Orth, J. Kauffman, and M. M. Davis. 2013. A database and meta-analysis of ecological responses to stream flow in the South Atlantic Region. Southeastern Naturalist 12:1–36.

Melcher, A. H., T. H. Bakken, T. Friedrich, F. Greimel, N. Humer, S. Schmutz, B. Zeiringer, and J. A. Webb. 2016. Drawing together multiple lines of evidence from assessment studies of hydropeaking pressures in impacted rivers. Freshwater Science 36(1):220–230.

Morsell, J. 2000. Long term studies of salmon in a regulted Alaskan stream. Proceedings of American Water Resources Association Spring Specialty Conference: Water Reserouces in Extreme Environments, Anchorage, Alaska:197–202.

Murchie, K. J., K. P. E. Hair, C. E. Pullen, T. D. Redpath, H. R. Stephens, and S. J. Cooke. 2008. Fish response to modified flow regimes in regulated rivers: research methods, effects and opportunities. River Research and Applications 24(2):197–217.

Naiman, R. J., J. J. Latterell, N. E. Pettit, and J. D. Olden. 2008. Flow variability and the biophysical vitality of river systems. Comptes Rendus Geoscience 340(9):629–643.

National Energy Board. 2018. Canada's Energy Future 2018: Energy Supply and Demand Projections to 2040. Government of Canada.

Nicol, E., J. R. Stevens, and S. Jobling. 2017. Riverine fish diversity varies according to geographical isolation and land use modification. Ecology and Evolution 7(19):7872–7883.

Nilsson, C., C. A. Reidy, M. Dynesius, and C. Revenga. 2005. Fragmentation and flow regulation of the world's large river systems. Science 308(5720):405–408.

Oberdorff, T., P. A. Tedesco, B. Hugueny, F. Leprieur, O. Beauchard, S. Brosse, and H. H. Dürr. 2011. Global and regional patterns in riverine fish species richness: A review. International Journal of Ecology 2011:e967631.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The natural flow regime. BioScience 47(11):769–784.

Poff, N. L., and J. K. H. Zimmerman. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology 55(1):194–205.

Pracheil, B. M., C. R. DeRolph, M. P. Schramm, and M. S. Bevelhimer. 2016. A fish-eye view of riverine hydropower systems: the current understanding of the biological response to turbine passage. Reviews in Fish Biology and Fisheries 26(2):153–167.

Pullin, A. S., and T. M. Knight. 2009. Doing more good than harm – Building an evidence-base for conservation and environmental management. Biological Conservation 142(5):931–934.

Pullin, A. S., and G. B. Stewart. 2006. Guidelines for systematic review in conservation and environmental management. Conservation Biology 20(6):1647–1656.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reid, A. J., A. K. Carlson, I. F. Creed, E. J. Eliason, P. A. Gell, P. T. J. Johnson, K. A. Kidd, T. J. MacCormack, J. D. Olden, S. J. Ormerod, J. P. Smol, W. W. Taylor, K. Tockner, J. C. Vermaire, D. Dudgeon, and S. J. Cooke. 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biological Reviews 94(3):849–873.

Reyjol, Y., P. Lim, F. Dauba, P. Baran, and A. Belaud. 2001. Role of temperature and flow regulation on the Salmoniform - Cypriniform transition. Archiv für Hydrobiologie 152(4):567–582.

Rohatgi, A. 2015. WebPlotDigitalizer: HTML5 based online tool to extract numerical data from plot images. https://automeris.io/WebPlotDigitizer

Rosenberg, M. S. 2005. The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. Evolution 59(2):464–468.

Rosenberg, M. S. 2013. Moment and least-squares based appraoches to meta-analytic inference. Pages 108–124 *in* J. Koricheva, J. Gurevitch, and K. Mengersen, editors. Handbook of Meta-analysis in Ecology and Evolution. Princeton University Press, Princeton, New Jersey.

Rosenthal, R. 1979. The file drawer problem and tolerance for null results. Psychological Bulletin 86(3):638–641.

Rosenthal, R. 1991. Meta-analytic Procedures for Social Research. Sage, London.

Rytwinski, T., S. J. Cooke, J. J. Taylor, D. G. Roche, P. A. Smith, G. W. Mitchell, K. E. Smokorowski, K. A. Prior, and J. R. Bennett. 2021. Acting in the face of evidentiary ambiguity, bias, and absence arising from systematic reviews in applied environmental science. Science of The Total Environment 775:145122.

Rytwinski, T., M. Harper, J. J. Taylor, J. R. Bennett, L. A. Donaldson, K. E. Smokorowski, K. Clarke, M. J. Bradford, H. Ghamry, J. D. Olden, D. Boisclair, and S. J. Cooke. 2020. What are the effects of flow-regime changes on fish productivity in temperate regions? A systematic map. Environmental Evidence 9(1):7.

Rytwinski, T., J. J. Taylor, J. R. Bennett, K. E. Smokorowski, and S. J. Cooke. 2017. What are the impacts of flow regime changes on fish productivity in temperate regions? A systematic map protocol. Environmental Evidence 6(1):13.

Schank, J. C., and T. J. Koehnle. 2009. Pseudoreplication is a pseudoproblem. Journal of Comparative Psychology 123(4):421–433.

Schumm, S. A. 2005. River Variability and Complexity. Cambridge University Press, New York.

Smokorowski, K. E., and R. G. Randall. 2017. Cautions on using the Before-After-Control-Impact design in environmental effects monitoring programs. FACETS 2:212-232.

Soininen, N., A. Belinskij, A. Vainikka, and H. Huuskonen. 2019. Bringing back ecological flows: migratory fish, hydropower and legal maladaptivity in the governance of Finnish rivers. Water International 44(3):321–336.

Studd, E. K., A. E. Bates, A. J. Bramburger, T. Fernandes, B. Hayden, H. A. L. Henry, M. M. Humphries, R. Martin, B. C. McMeans, E. R. D. Moise, A. M. O'Sullivan, S. Sharma, B.

J. Sinclair, A. O. Sutton, P. H. Templer, and S. J. Cooke. In press. Nine maxims for the ecology of cold-climate winters. BioScience 00:000–000.

Su, G., M. Logez, J. Xu, S. Tao, S. Villéger, and S. Brosse. 2021. Human impacts on global freshwater fish biodiversity. Science 371(6531):835–838.

Sullivan, R. M., and J. P. Hileman. 2018. Annual and seasonal variation, relative abundance, and effects of managed flows on timing of migration in Brown Trout (*Salmo trutta*) in the upper Trinity River 104(3):99–128.

Tanner-Smith, E. E., and E. Tipton. 2014. Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and spss. Research Synthesis Methods 5(1):13–30.

Tonkin, J. D., N. L. Poff, N. R. Bond, A. Horne, D. M. Merritt, L. V. Reynolds, J. D. Olden, A. Ruhi, and D. A. Lytle. 2019. Prepare river ecosystems for an uncertain future. Nature 570(7761):301–303.

Travnichek, V. H., M. B. Bain, and M. J. Maceina. 1995. Recovery of a warmwater fish assemblage after the initiation of a minimum-flow release downstream from a hydroelectric dam. Transactions of the American Fisheries Society 124(6):836–844.

Turgeon, K., C. Turpin, and I. Gregory-Eaves. 2019. Dams have varying impacts on fish communities across latitudes: a quantitative synthesis. Ecology Letters 22(9):1501–1516.

Ugedal, O., T. F. Næsje, E. B. Thorstad, T. Forseth, L. M. Saksgård, and T. G. Heggberget. 2008. Twenty years of hydropower regulation in the River Alta: long-term changes in abundance of juvenile and adult Atlantic salmon. Hydrobiologia 609(1):9–23.

U.S. Department of Energy. 2016. Hydropower Vision: A New Chapter for America's 1st Renewable Electricity Source.

Viechtbauer, W. 2010. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 36(1):1–48.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human domination of Earth's ecosystems. Science 277(5325):494–499.

Vittinghoff, E., S. C. Shiboski, D. V. Glidden, and C. E. McCulloch. 2005. Regression Methods in Biostatistics. Springer, New York.

Webb, J. A., K. A. Miller, E. L. King, S. C. de Little, M. J. Stewardson, J. K. H. Zimmerman, and N. L. Poff. 2013. Squeezing the most out of existing literature: A systematic re-analysis of published evidence on ecological responses to altered flows. Freshwater Biology 58(12):2439–2451.

Williams, G. P., and M. G. Wolman. 1987. Downstream effects of dams on alluvial rivers. U.S. Department of the Interior, 1286, United States Government Printing Office, Washington.

Winemiller, K. O., P. B. McIntyre, L. Castello, E. Fluet-Chouinard, T. Giarrizzo, S. Nam, I. G. Baird, W. Darwall, N. K. Lujan, I. Harrison, M. L. J. Stiassny, R. a. M. Silvano, D. B. Fitzgerald, F. M. Pelicice, A. A. Agostinho, L. C. Gomes, J. S. Albert, E. Baran, M. Petrere, C. Zarfl, M. Mulligan, J. P. Sullivan, C. C. Arantes, L. M. Sousa, A. A. Koning, D. J. Hoeinghaus, M. Sabaj, J. G. Lundberg, J. Armbruster, M. L. Thieme, P. Petry, J. Zuanon, G. T. Vilara, J. Snoeks, C. Ou, W. Rainboth, C. S. Pavanelli, A. Akama, A. van Soesbergen, and L. Sáenz. 2016. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. Science 351(6269):128–129.

Young, P. S., J. J. Cech, and L. C. Thompson. 2011. Hydropower-related pulsed-flow impacts on stream fishes: a brief review, conceptual model, knowledge gaps, and research needs. Reviews in Fish Biology and Fisheries 21(4):713–731.

Zarfl, C., A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner. 2015. A global boom in hydropower dam construction. Aquatic Sciences 77(1):161–170.

Zwetsloot, P.-P., M. Van Der Naald, E. S. Sena, D. W. Howells, J. IntHout, J. A. De Groot, S. A. Chamuleau, M. R. MacLeod, and K. E. Wever. 2017. Standardized mean differences cause funnel plot distortion in publication bias assessments. eLife 6:e24260.