ORIGINAL RESEARCH

Ecology and Evolution
Open Access

WILEY

Ignoring non-English-language studies may bias ecological meta-analyses

Ko Konno[1] | Munemitsu Akasaka[2,3] | Chieko Koshida[4] | Naoki Katayama[5] | Noriyuki Osada[6] | Rebecca Spake[7] | Tatsuya Amano[3,8,9]

[1]School of Natural Sciences, Bangor University, Gwynedd, UK

[2]Institute of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Japan

[3]Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Fuchu, Japan

[4]Plant Chemical Ecology, Technische Universität Darmstadt, Darmstadt, Germany

[5]Biodiversity Division, Institute for Agro-Environmental Sciences NARO, Tsukuba-shi, Japan

[6]Laboratory of Plant Conservation Science, Faculty of Agriculture, Meijo University, Nagoya, Japan

[7]School of Geography and Environmental Science, University of Southampton, Southampton, UK

[8]School of Biological Sciences, The University of Queensland, Brisbane, Qld, Australia

[9]Centre for Biodiversity and Conservation Science, The University of Queensland, Brisbane, Qld, Australia

Correspondence
Tatsuya Amano, School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia.
Email: t.amano@uq.edu.au

Abstract

Meta-analysis plays a crucial role in syntheses of quantitative evidence in ecology and biodiversity conservation. The reliability of estimates in meta-analyses strongly depends on unbiased sampling of primary studies. Although earlier studies have explored potential biases in ecological meta-analyses, biases in reported statistical results and associated study characteristics published in different languages have never been tested in environmental sciences. We address this knowledge gap by systematically searching published meta-analyses and comparing effect-size estimates between English- and Japanese-language studies included in existing meta-analyses. Of the 40 published ecological meta-analysis articles authored by those affiliated to Japanese institutions, we find that three meta-analysis articles searched for studies in the two languages and involved sufficient numbers of English- and Japanese-language studies, resulting in four eligible meta-analyses (i.e., four meta-analyses conducted in the three meta-analysis articles). In two of the four, effect sizes differ significantly between the English- and Japanese-language studies included in the meta-analyses, causing considerable changes in overall mean effect sizes and even their direction when Japanese-language studies are excluded. The observed differences in effect sizes are likely attributable to systematic differences in reported statistical results and associated study characteristics, particularly taxa and ecosystems, between English- and Japanese-language studies. Despite being based on a small sample size, our findings suggest that ignoring non-English-language studies may bias outcomes of ecological meta-analyses, due to systematic differences in study characteristics and effect-size estimates between English- and non-English languages. We provide a list of actions that meta-analysts could take in the future to reduce the risk of language bias.

KEYWORDS

language barriers, publication bias, reporting biases, risk of bias, systematic review

---

1 | INTRODUCTION

Global environmental change threatens ecosystems and biodiversity around the world (Ceballos et al., 2015; Díaz et al., 2019; WWF, 2018). A sound understanding of ecosystem responses to environmental drivers and human activities is therefore urgently required to inform policy and practice to mitigate against adverse ecological change (Sutherland, Pullin, Dolman, & Knight, 2004). It is increasingly demanded that this understanding draws on rigorous scientific evidence bases, best formed through the unbiased and systematic collation, appraisal, and meta-analysis of primary empirical research (Pullin, 2012; Sutherland et al., 2004). Meta-analysis can provide a powerful set of tools for summarizing the results of multiple studies, quantifying the variation in results among studies, and evaluating whether hypotheses are supported by the assemblage of existing studies (Gurevitch, Koricheva, Nakagawa, & Stewart, 2018; Koricheva, Gurevitch, & Mengersen, 2013).

A common criticism of meta-analysis is the ignorance of potential biases during the search and selection of studies to be quantitatively synthesized (Borenstein, Hedges, Higgins, & Rothstein, 2009). Indeed, for a meta-analysis to obtain a robust estimate of an overall true effect size, a random subset of all relevant primary studies should be included in the analysis. For example, because the nature and direction of a study's results can affect its likelihood of publication (publication bias: Bayliss & Beyer, 2015; CEE, 2018; Higgins & Green, 2011), the omission of unpublished data and gray literature may result in a biased sample of primary studies that give rise to an overestimated overall effect size (McAuley, Pham, Tugwell, & Moher, 2000; Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008). Publication bias and its consequences for meta-analysis are widely recognized, and mitigation measures to minimize this bias exist (Bayliss & Beyer, 2015; CEE, 2018). A much more overlooked bias in evidence synthesis is language bias, wherein the nature and direction of a study's results can affect the chosen language of its publication (Egger et al., 1997; Grégoire, Derderian, & Le Lorier, 1995; Higgins & Green, 2011; Juni et al., 2002). Omitting studies published in languages other than English, a common practice in meta-analysis, could therefore also lead to a biased sample of primary studies. However, the prevalence and importance of language bias in ecological meta-analyses have never been assessed to date (Livoreil et al., 2017). This is concerning, given that conclusions derived from biased meta-analyses could lead to wasted resources if management actions are ineffectively prescribed or, worse still, may lead to unexpected or even perverse outcomes.

Previous studies of language bias in medical science have revealed differences in statistical results between publication languages (Egger et al., 1997; Grégoire et al., 1995; Juni et al., 2002). Referred to as "English-language bias" (Egger et al., 1997) or "Tower of Babel" bias (Grégoire et al., 1995), it has been shown that positive or statistically significant results are more likely to be published in English than other languages (*language bias in statistical results* in Figure 1). This focus on statistical results is presumably because medical meta-analysts are typically concerned with estimating the overall effects of treatments on a single species (i.e., *Homo sapiens*)

under controlled conditions (e.g., the effectiveness of a drug at reducing symptoms of a disease). In contrast, ecological meta-analyses are typically interested in variation among effect sizes and attributing this variation to meaningful covariates that vary among studies, such as species biogeographical contexts and intervention intensity (i.e., effect modification). Ecological meta-analyses thus typically combine heterogeneous studies on a wide range of organisms and ecosystems (Gurevitch et al., 2018; Koricheva et al., 2013). Doing so may give rise to another type of language bias, if studies with particular characteristics are more likely to be published in non-English languages because, for instance, they are deemed unsuitable for international journals. For example, studies conducted on particular ecosystems, at particular intervention intensities, or conducted by local practitioners who do not speak English, could be systematically omitted from meta-analyses (*language bias in study characteristics* in Figure 1). Considering that up to 36% of scientific studies on biodiversity conservation is published in languages other than English (Amano, González-Varo, & Sutherland, 2016), and that non-English studies are typically omitted from ecological meta-analyses, an assessment of the impacts of language bias on ecological inferences drawn from meta-analyses is urgently needed.

Here, we address this knowledge gap and assess the risk of language bias on outcomes of meta-analyses. We searched for published peer-reviewed meta-analysis articles that analyzed sufficient numbers of both English- and Japanese-language peer-reviewed studies (10 or more effect-size estimates in each language). The Japanese language was chosen not only for a practical reason (most of the authors are native Japanese-language speakers) but also because the Japanese language is one of the major non-English languages for scientific documentation in biodiversity conservation (Amano et al., 2016). We first tested the differences in effect-size estimates between English- and Japanese-language studies included in eligible meta-analyses, to quantify the impacts of excluding the Japanese-language studies on the overall mean effect sizes. To investigate the possible causes of differences in effect-size estimates between the languages, we then tested for language bias in study characteristics by investigating between-language differences in study characteristics that were deemed important in modifying the effects in the original meta-analyses (i.e., potential effect modifiers). Next, we quantified differences in effect-size estimates between the languages after controlling for the revealed differences in the study characteristics. Finally, we discuss the processes through which language bias may arise and propose guidelines for incorporating non-English-language studies in ecological meta-analyses.

2 | MATERIALS AND METHODS

2.1 | Systematic literature search

We searched for ecological meta-analysis articles including both English-language and Japanese-language studies. We performed searches in Web of Science Core Collection (https://webofknowl
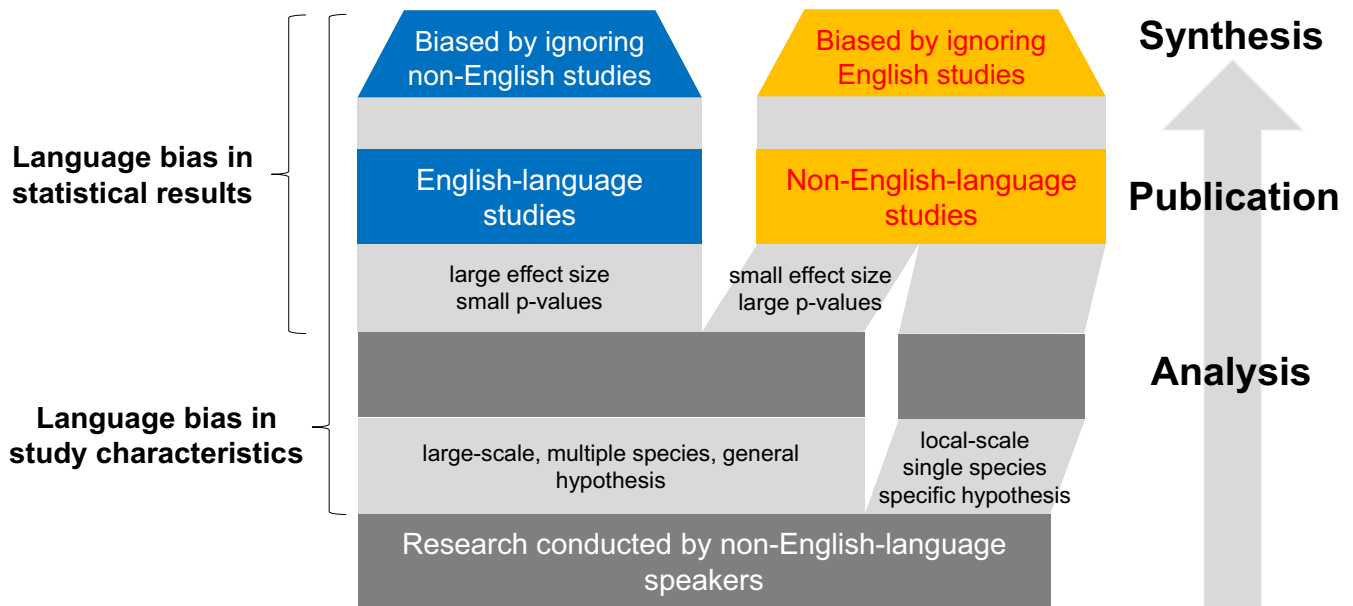
FIGURE 1 The fate of ignoring studies published in relevant language(s). Studies providing certain information (e.g., local-scale studies focusing on specific hypotheses on a single species) may be more likely to be published in non-English languages (*language bias in study characteristics*) because, for example, those studies tend to be conducted by local practitioners or they are often not of great interest from an international perspective. After the analysis, statistically significant or positive results may be more likely to be published in higher-impact, English-language journals (*language bias in statistical results*)

edge.com/), CAB Direct (https://www.cabdirect.org/), and Wiley Online Library (https://onlinelibrary.wiley.com/) using English-language search strings via Bangor University institutional access, and in CiNii (https://ci.nii.ac.jp/), which is the largest and most comprehensive database in Japan, using a Japanese-language search string with no subscription (Appendix S1). We also performed Web-based searches on Google Scholar (in English: https://scholar.google.com/; in Japanese: https://scholar.google.co.jp/) using an English-language search string and a Japanese-language search string (Appendix S1). Each search string contains meta-analy* OR "meta analy*." We tailored a search string for each bibliographic platform, database, and Web-based search engine (Appendix S1). We used translated version of the terms for the searches in Japanese (AppendixS1). All searches were conducted in Gwynedd, Wales, United Kingdom, on 19, 24, and 25 July 2018, and the searches were updated on 29 March 2019 (see exact time and date for each search in AppendixS1 https://refworks.proquest.com/); Mendeley (https://www.mendeley.com/); and EndNote Basic (https://endnote.com). Prior to the full searches and result retrieval, we had developed and predefined searches and result retrieval through a pilot test (AppendixS2 and AppendixS3). For the pilot searches, we adopted search strings developed by O'Leary et al. (2016).

2.2 | Eligibility screening

Our search strings retrieved 1,504 unique articles. Duplicates were removed using Mendeley's "Check for Duplicates" tool and manually. We then screened articles according to titles and abstract, followed by full texts. We screened these articles to obtain meta-analysis articles on ecological or evolutionary topics, suitable for assessing possible language biases. We included meta-analysis articles that expressed the outcome of multiple studies on a common scale, through the calculation of an "effect size" for each study, which represents the magnitude of a difference between control and treatment means (e.g., log response ratio, standardized mean difference). We included meta-analysis articles that aimed to quantitatively combine effect sizes to yield an overall estimate, or attribute variation in effect sizes to meaningful covariates using meta-regression. Note that a meta-analysis article can include multiple meta-analyses. To identify meta-analysis articles authored by individuals capable of searching literature in both English and Japanese, we included only meta-analysis articles conducted by research teams with at least one author affiliated with a Japanese institution. Finally, the meta-analyses had to include 10 or more effect-size estimates published in both English and Japanese languages (AppendixS4). Articles meeting all of these eligibility criteria were included in our analysis. We used a modified version of the ROSES Flow Diagram for Systematic Review for reporting the number of articles retrieved at each screening stage (Haddaway, Macura, Whaley, & Pullin, 2017, 2018) (AppendixS5). We do not report a critical appraisal component of the diagram, because we did not conduct critical appraisal of the meta-analysis articles.

2.3 | Data selection

We extracted effect-size estimates provided by four eligible meta-analyses published in three meta-analysis articles by Koshida and

Katayama (2018), Osada et al. (2013), and Spake et al. (2019). Koshida and Katayama (2018) included 64 effect-size estimates from 35 studies on the effects of rice-field abandonment on biodiversity. We excluded six effect-size estimates from four studies from non-peer-reviewed reports to distinguish the effect of language bias from publication bias. The meta-analysis of Osada et al. (2013) compared the effect of light on leaf life span of four types of plants: deciduous herbaceous plants; evergreen herbaceous plants; deciduous woody plants; and evergreen woody plants. Only data of evergreen woody plants had sufficient numbers of both English- and Japanese-language studies. Eight unpublished studies were also excluded. Data provided by Spake et al. (2019) included effect sizes representing forestry impacts on the species richness and abundance of three population groups: ground-layer plants; saplings and seedlings; and invertebrates. Sufficient numbers of both English- and Japanese-language studies were available for meta-analyses of thinning impacts on the abundance of ground-layer plants, and saplings and seedlings. Thus, we excluded data of the effect of thinning on species richness and on abundance of invertebrates.

As a result, we used 58 effect-size estimates: 11 from six English-language studies and 47 from 25 Japanese-language studies from Koshida and Katayama (2018) ("rice-field meta-analysis" from hereon; Table 1 and Appendix S6). We also used 134 effect-size estimates: 100 from 13 English-language studies and 34 from two Japanese-language studies from Osada et al. (2013) ("leaf life span meta-analysis" from hereon; Table 1 and Appendix S7). From Spake et al. (2019), we used 65 effect-size estimates: 41 from six English-language studies and 24 from three Japanese-language studies on the effect of thinning on abundance of ground-layer plants ("plant forestry meta-analysis" from hereon; Table 1 and Appendix S8), and 41 effect-size estimates: 26 from six English-language studies and 15 from four Japanese-language studies on the effect of thinning on abundance of saplings and

seedlings ("sapling forestry meta-analysis" from hereon; Table 1 and Appendix S9).

2.4 | Data analysis

2.4.1 | Effect-size difference between languages

We first tested homogeneity of variance and normality of effect-size estimates (log response ratio and life span ratio) using Levene's test and two-sample Kolmogorov–Smirnov test, respectively. We then tested differences in effect-size estimates between English-language studies and Japanese-language studies, using a two-sample t test or Welch two-sample t test (Ruxton, 2006) where the assumption of homogeneity of variance was not met. While these analyses assume independence between effect-size estimates, some studies included in each meta-analysis have multiple effect-size estimates. Although some of those effect-size estimates (e.g., those estimated at different locations) can be assumed as independent comparisons, others may not, for example, by sharing common control groups. We could not however include study as a random effect (a common solution in meta-analyses, e.g., see Spake et al., 2019), because (a) an unbalanced number of effect-size estimates in each study (ranging from 1 to 36 in our analyses) can lead to unstable parameter estimates (Harrison et al., 2018), and (b) language is a study-level variable and can therefore be confounded with the random effect of each study. Therefore, our statistical tests may be vulnerable to increased type I errors. However, even if effect-size estimates are not independent to each other, the estimation of a mean effect size is not affected by this (Borenstein et al., 2009), and thus, mean effect sizes in each language shown in Figures 2 and 6 should be reliable.

Meta-analyses typically estimate precision-weighted mean effect sizes, with each effect size weighted by the inverse of its variance (in addition to between-study variance). Studies that omit the

TABLE 1 Details of eligible meta-analyses, the number of studies, and the number of effect-size estimates (total, in English, and in Japanese) included in each meta-analysis

Meta-analysis articles	Meta-analyses	Number of studies			Number of effect-size estimates		
		Total	English	Japanese	Total	English	Japanese
Koshida and Katayama (2018)	Rice-field meta-analysis	31	6 (19%)	25 (81%)	58	11 (19%)	47 (81%)
Osada et al. (2013)	Leaf life span meta-analysis	15	13 (87%)	2 (13%)	134	100 (75%)	34 (25%)
Spake et al. (2019)	Plant forestry meta-analysis	9	6 (67%)	3 (33%)	65	41 (63%)	24 (37%)
	Sapling forestry meta-analysis	10	6 (60%)	4 (40%)	41	26 (63%)	15 (37%)
	Total	62	29 (47%)	33 (53%)	298	178 (60%)	120 (40%)

Note: The two meta-analyses in Spake et al. (2019) used some studies in common, and thus, the total number of studies does not equate to the sum of studies used in each meta-analysis. Meta-analysis article: a published article (paper) that conducted at least one relevant meta-analysis. Meta-analysis: a statistical analysis of multiple effect-size estimates measuring the effect of an intervention on a distinct group of subjects. Study: a paper included in a meta-analysis, providing at least one effect-size estimate. Effect-size estimate: effect sizes estimated from data published in original primary studies and used in a meta-analysis.

FIGURE 2 Differences in mean effect sizes between English- (blue) and Japanese-language studies (red). (a) Koshida and Katayama (2018) showing the effects of rice-field abandonment on biodiversity (rice-field meta-analysis). (b) Osada et al. (2013) showing the effect of light on plants' leaf life span (leaf life span meta-analysis). (c) Spake et al. (2019) showing the effect of thinning on ground-layer plant abundance (plant forestry meta-analysis). (d) Spake et al. (2019) showing the effect of thinning on sapling and seedling abundance (sapling forestry meta-analysis). The number of effect-size estimates in each language is also shown. The error bars show 95% confidence intervals. Diamonds and dotted lines show mean effect sizes estimated from all studies pooled (i.e., English- + Japanese-language studies)
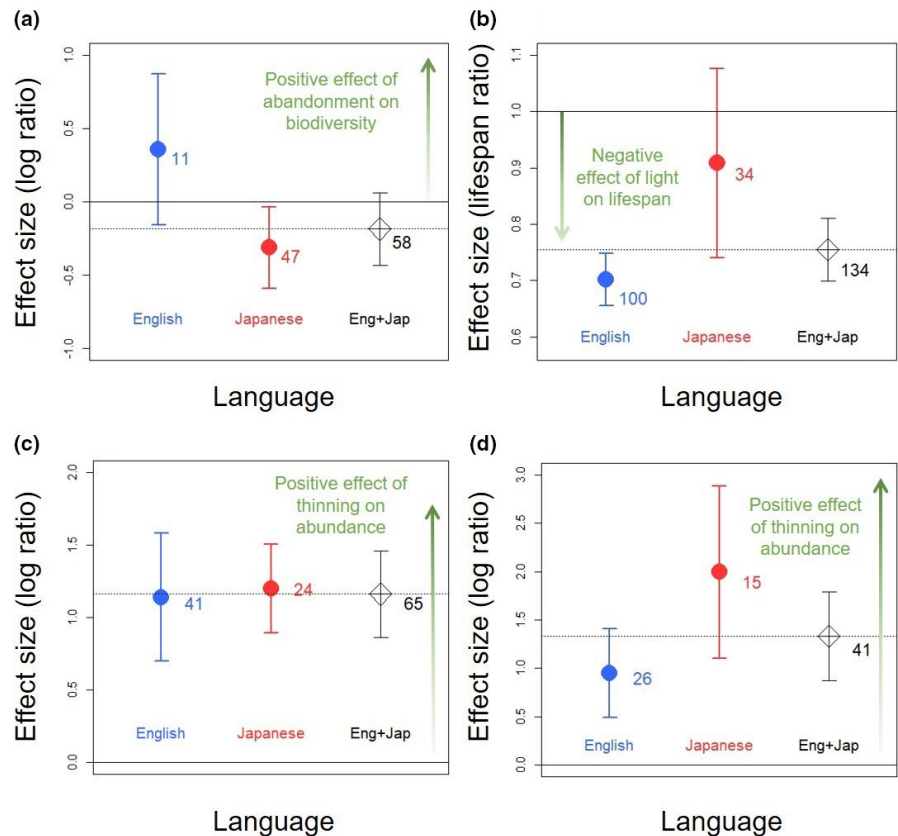
required information to estimate effect-size variance may thus be excluded (Borenstein et al., 2009). Therefore, as a sensitivity analysis, we repeated the above analyses but on a subset of effect-size estimates from studies that reported corresponding variance measures. We did not perform this sensitivity analysis for Osada et al. (2013) because the original meta-analysis did not provide standard deviations of effect-size estimates.

2.4.2 | Language bias in study characteristics

We tested associations between the languages and study characteristics using two-way chi-squared tests and Mann–Whitney U tests. For Koshida and Katayama (2018), taxa, management types, landscape types, soil types, and outcomes measured were analyzed. For Osada et al. (2013), measurement conditions, study countries, and plant families were analyzed. Study countries were tested because Japanese-language studies are not necessarily conducted in Japan. In the case of Spake et al. (2019), intervention intensity and stand age were analyzed. We chose these variables because the original meta-analyses treated the variables as potential effect modifiers and they were available for analyses. Note that although strictly speaking these are the characteristics of each effect-size estimate, many of those characteristics are usually determined at the study level; hence, we used the term "study characteristics" instead of "characteristics of effect-size estimates." The statistical tests were conducted in R version 3.5.0 (R Core Team, 2018).

2.4.3 | Language bias in statistical results

We used two approaches to assess whether effect-size estimates differed between languages after controlling differences in study characteristics. First, for each meta-analysis, we fitted linear models with effect-size estimates as the response variable, with explanatory variables including publication language in addition to factors that had a significant association with language (see above) as fixed factors, then compared two models: with and without the fixed factor language. Second, we fitted linear mixed models with effect-size estimates as the response variable, publication language as the fixed factor, and factors that had a significant association with language as random factors, and tested the significance level of the fixed factors using likelihood-ratio tests with the reduced models (Quinn & Keough, 2002). We ran the linear mixed models and performed likelihood-ratio tests using lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.5.0 (R Core Team, 2018).

3 | RESULTS

3.1 | Searches and screening

Of the 1,504 unique articles retrieved by our search strings, 40 articles met our inclusion criteria as ecological meta-analyses conducted by at least one author affiliated with a Japanese institution. These comprised meta-analyses from a wide range of subdisciplines,

including forestry, phenology, agriculture, and ecosystem services. Only three published meta-analysis articles searched for studies published in both English and Japanese languages. Thirty-six English-language meta-analysis articles included studies published only in English, while one Japanese-language meta-analysis article synthesized only Japanese-language studies (listed as "Evidence base" for the reason for exclusion in AppendixS10). From the three articles meeting our criteria, four separate meta-analyses had sufficient data to examine potential language bias effects (Table 1; also see Section 2.3 Data Selection in Methods). Japanese-language studies constituted 81% of the effect-size estimates from the rice-field meta-analysis, 25% from the leaf life span meta-analysis, and 37% each from the two forestry meta-analyses (Table 1).

3.2 | Effect-size differences between languages

Effect sizes representing the effect of rice-field abandonment on biodiversity differed between English- and Japanese-language studies from the rice-field meta-analysis (Table 2, Figure 2a). Both the magnitude and direction of the mean effect sizes differed among languages. A positive mean effect was estimated from English-language-only studies, whereas a negative mean effect was estimated from Japanese-language-only studies, and from all studies pooled (Figure 2a). This result did not change when only a subset of effect-size estimates that had reported standard deviations were analyzed (i.e., those that allow weighted meta-analysis: AppendixS11.

Similarly, effect sizes for the leaf life span meta-analysis differed between the languages (Table 2, Figure 2b). . The mean effect size estimated from English-language-only studies showed a more strongly negative effect of light on leaf life span than mean effect sizes estimated from both Japanese-language studies and all studies, by 23% and 7%, respectively (Figure 2b).

In contrast, effect sizes did not differ significantly between English-language and Japanese-language studies from the plant forestry and sapling forestry meta-analyses (Table 2). However, mean effect sizes from English-language-only studies in the sapling forestry

meta-analysis were 52% and 29% smaller than those estimated from Japanese-language-only studies and from all studies, respectively (Figure 2d). Results differed for comparisons based on the subset of effect-size estimates that were associated with standard deviations (i.e., those that allow weighted meta-analysis). Effect sizes differed significantly between languages in the plant forestry meta-analysis. The mean effect size estimated from Japanese-language-only studies was 318% more positive than the mean effect from English-language-only studies in the plant forestry meta-analysis ($t = -2.85$; $df = 29$; $p = .008$; AppendixS12). The difference remained nonsignificant in the sapling forestry meta-analysis ($t = -0.98$; $df = 27$; $p = .33$).

3.3 | Language bias in study characteristics

For the rice-field meta-analysis, although all of the included studies were conducted in Japan, the proportion of Japanese-language studies varied significantly among taxa ($x^2 = 25.07$; $df = 3$; $p < .001$) and landscape types ($x^2 = 14.38$; $df = 1$; $p < .001$; Figure 3). Almost all studies on amphibians, fish, and plants and in complex landscapes were those published in Japanese (Figure 3).

For the leaf life span meta-analysis, the proportion of Japanese-language studies differed significantly among measurement conditions ($x^2 = 5.05$; $df = 1$; $p = .02$), plant families ($x^2 = 109.07$; $df = 27$; $p < .001$), and not surprisingly study countries ($x^2 = 103.34$; $df = 7$; $p < .001$; Figure 4). Most studies with experimental measurements were those published in English, while all studies on four families, Aquifoliaceae, Oleaceae, Rosaceae, and Garryaceae, were those published in Japanese (Figure 4).

In the forestry meta-analyses, the moderating influences of forest thinning intensity (volume removed, %) and stand age on effect sizes were analyzed (Spake et al., 2019). Despite the fact that all studies were conducted in Japan, thinning intensities were higher for studies published in Japanese than English for both the plant forestry meta-analysis ($U_{lower} = 350$; n1 = 41; n2 = 24; $p = .049$) and the sapling forestry meta-analysis ($U_{lower} = 109$; n1 = 26; n2 = 15; $p = .02$; Figure 5a). Stand age was younger in Japanese-language

Meta-analysis	Levene's test for homogeneity of variance		Two-sample Kolmogorov–Smirnov test for normality		Two-sample t test for effect-size differences between languages	
	F (df)	p	D	p	t (df)	p
Rice-field meta-analysis	0.13 (1, 56)	.72	0.44	.06	2.18 (56)	**.03**
Leaf life span meta-analysis	4.55 (1, 132)	**.03**	0.27	.08	−2.40 (38.42)	**.02**
Plant forestry meta-analysis	1.68 (1, 63)	.20	0.29	.12	−0.19 (63)	.85
Sapling forestry meta-analysis	6.07 (1, 39)	**.02**	0.36	.17	−2.03 (21.62)	.05

TABLE 2 Results of statistical tests for homogeneity of variance, normality, and differences in effect sizes between English- and Japanese-language studies

Note: Statistically significant results are in bold. Welch two-sample t test was used where the assumption of homogeneity of variance was not met.
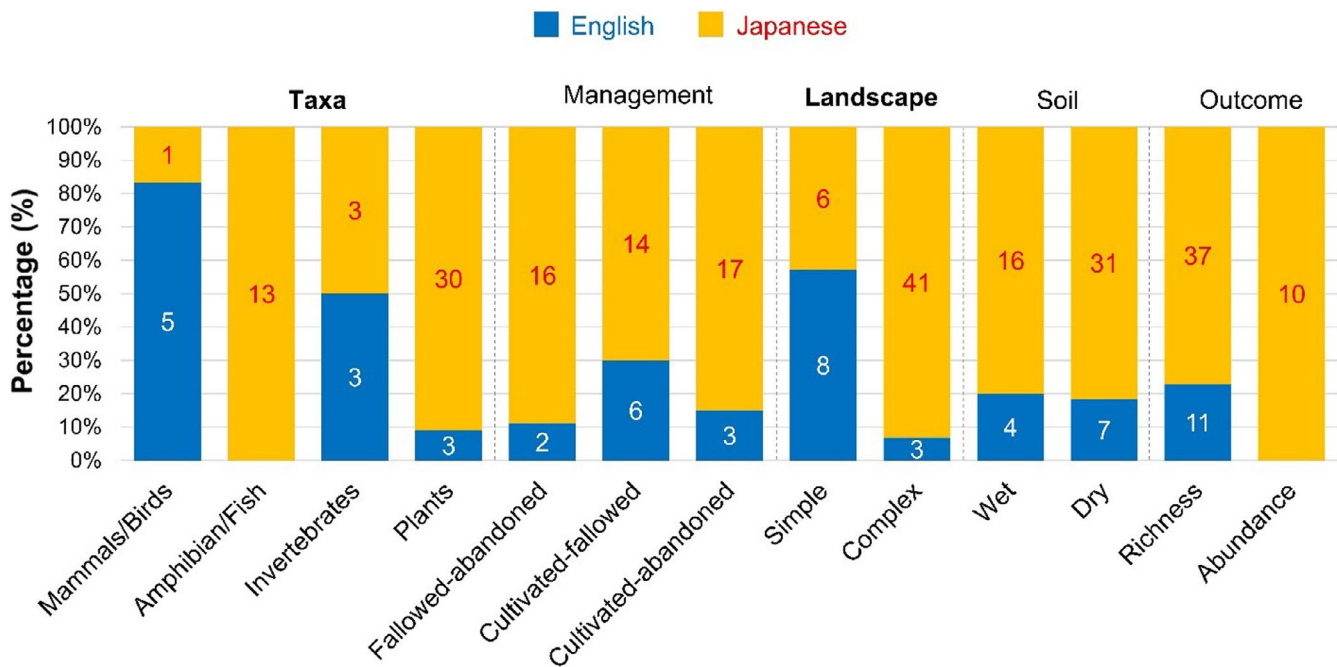
FIGURE 3 Differences in study characteristics between English- (blue) and Japanese-language studies (orange) for the rice-field meta-analysis (Koshida & Katayama, 2018). Characteristics with a significant difference between the languages are in bold. The number of effect-size estimates in each language is also shown in each bar

studies for the plant forestry meta-analysis (U_{lower} = 646; n1 = 41; n2 = 24; p = .04), but it did not differ significantly between the languages for the sapling forestry meta-analysis (U_{lower} = 168; n1 = 26; n2 = 15; p = .47; Figure 5b). This difference remained when reanalyzed for the subset of effect-size estimates associated with sample sizes and standard deviations for the plant forestry meta-analysis (U_{lower} = 54.5; n1 = 13; n2 = 18; p = .01); however, stand age did not differ significantly between the languages (U_{lower} = 131; n1 = 13; n2 = 18; p = .58).

3.4 | Language bias in statistical results

We next tested whether differences in effect sizes among languages remained after controlling for differences in study characteristics between the English-language and Japanese-language studies. In the rice-field meta-analysis, effect sizes did not differ significantly between the languages when the effects of taxa and landscape types were controlled for (Table 3, Figure 6a,b). Thus, the differences in effect sizes between the languages shown in Figure 2a seem to be largely attributable to the differences in study characteristics, together with the differences in effect sizes between taxa and landscape types (Figure 6a,b).

In contrast, effect sizes from the leaf life span meta-analysis still differed among languages after controlling for the effects of measurement conditions, plant families, and study countries (Table 3, Figure 6c). English-language studies tended to report larger effect sizes than the Japanese-language studies (Figure 6c), even for

studies conducted only in Japan or those with field measurements, indicating a systematic difference in reported statistical results.

Effect-size estimates associated with standard deviations from the plant forestry meta-analysis differed among languages after controlling for the difference in forest thinning intensity (Table 3, Figure 6d). This again suggests a systematic difference in reported statistical results between languages, but in this case, the Japanese-language studies tended to report larger effect sizes than the English-language studies (Figure 6d).

4 | DISCUSSION

Our reanalysis of four published meta-analyses has revealed differences in mean effect sizes from studies published in English and Japanese. Specifically, two out of the four multiple-language ecological meta-analyses (and in another, when focusing only on effect-size estimates associated with standard deviations) had effect sizes that differed by language. Although most of the ecological meta-analyses identified in our initial screening neither searched for nor included studies published in multiple languages, Japanese-language studies constituted a high proportion of effect-size estimates included in the four eligible meta-analyses (Table 1). These results suggest that English-language studies do not necessarily comprise a random subset of the global literature, and thus, ignoring non-English-language studies in ecological meta-analyses may lead to biased estimates of overall mean effect sizes, and biased inferences about ecological effects. This is a serious concern, given that meta-analyses are often
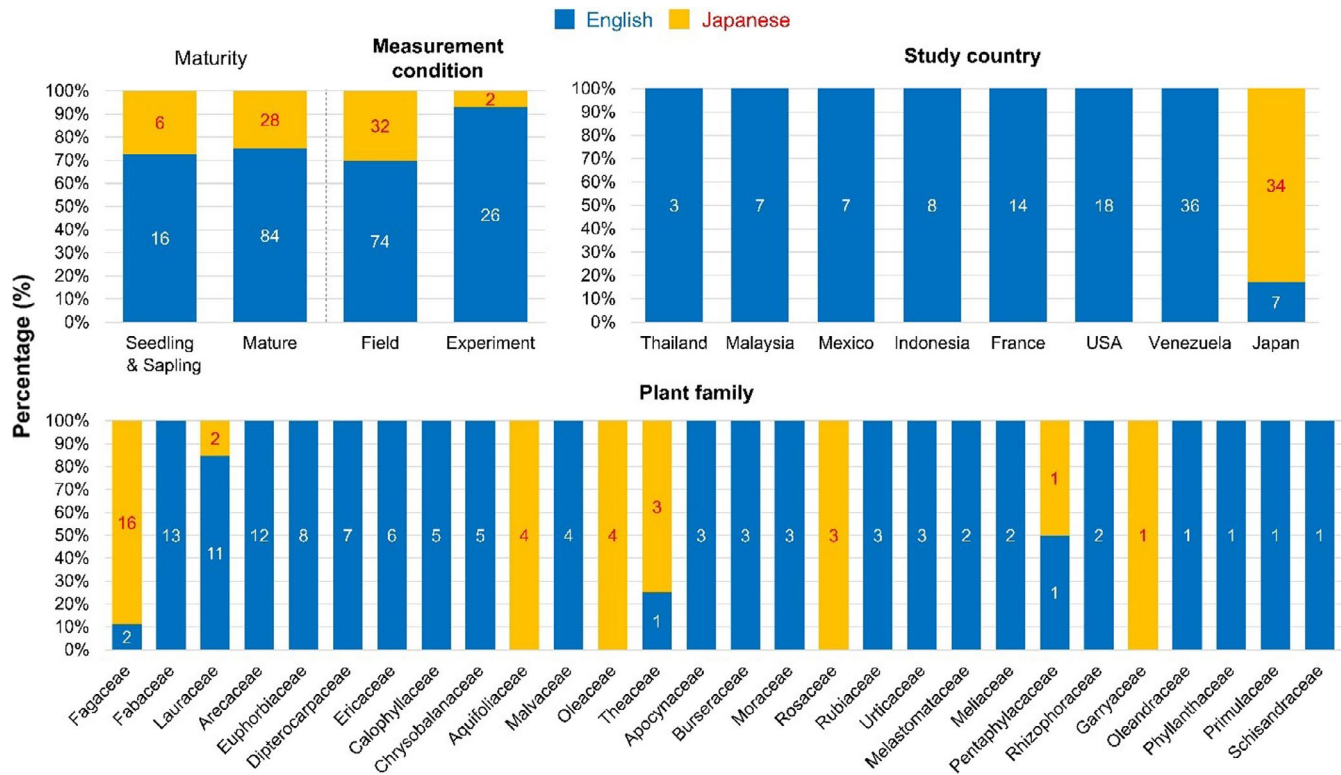
FIGURE 4 Differences in study characteristics between English- (blue) and Japanese-language studies (orange) for the leaf life span meta-analysis (Osada et al., 2013). Characteristics with a significant difference between the languages are in bold. The number of effect-size estimates in each language is also shown in each bar
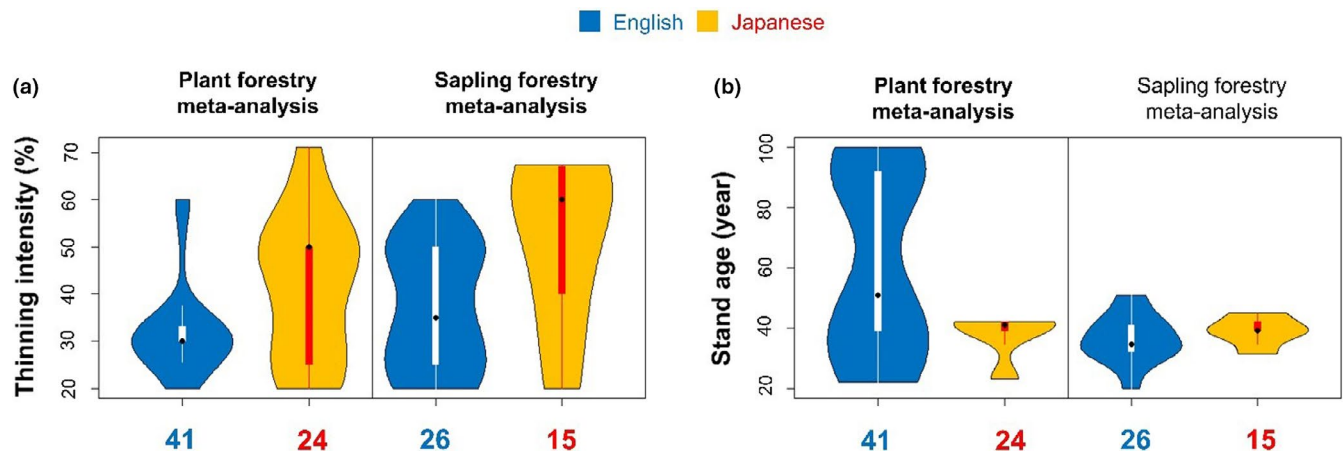

FIGURE 5 Differences in two study characteristics, (a) thinning intensity and (b) stand age, between English- (blue) and Japanese-language studies (orange) for the plant and sapling forestry meta-analyses (Spake et al., 2019). Black circles show medians, squares show interquartile ranges, and outer lines show ranges. The number of effect-size estimates in each language is also shown below each panel

used to inform decision-making in conservation policy and practice across a range of contexts (Gurevitch et al., 2018).

Our results also illustrate that language bias in study characteristics could lead to biased estimates of overall mean effect sizes, if non-English-language studies are ignored. This is a potential threat to ecological meta-analyses, which typically synthesize heterogeneous studies on multiple species, conducted at multiple study scales (temporal, spatial) and with different methods (e.g., observational,

experimental) (Spake & Doncaster, 2017). It is especially hard, or even impossible, to address the effect of language bias in study characteristics on overall mean effect sizes, when none or only a few effect-size estimates are available in English for some study characteristics (e.g., effect-size estimates on amphibians, fish, and plants in Koshida & Katayama, 2018). As a substantial amount of scientific literature on specific ecosystems and endemic species seems to be published in non-English languages (Amano et al., 2016), omitting

TABLE 3 Comparisons of models for testing effect-size differences between languages, with and without language as a fixed factor

Meta-analysis	Fixed factor(s)	Random factor(s)	Comparison of models with and without language as the fixed factor		
			F	x^2	p
Rice-field meta-analysis	Language + Taxa		0.26		.62
	Language + Landscape		0.14		.71
	Language	Taxa		1.97	.16
	Language	Landscape		0.58	.45
	Language	Taxa + Landscape		0.17	.68
Leaf life span meta-analysis	Language + Measurement condition		12.64		**.0005**
	Language + Plant family		12.76		**.0005**
	Language	Measurement condition		10.68	**.001**
	Language	Plant family		13.59	**.0002**
	Language	Study country		16.00	**<.0001**
	Language	Measurement condition + Plant family + Study country		16.10	**<.0001**
Plant forestry meta-analysis	Language + Thinning intensity		4.23		**.049**
	Language	Thinning intensity		12.06	**.0005**

Note: Statistically significant results (in bold) indicate that effect sizes differ between English- and Japanese-language studies even after controlling for the relevant fixed or random factor(s).

non-English-language studies might result in biased samples of study characteristics, potentially leading to invalid inferences. The general risk of language bias in study characteristics might thus be higher for ecology than for medical sciences.

We discuss four principal reasons that might give rise to language bias in study characteristics for Japan. First, there are well-established national and local societies in Japan that publish their own journals, such as the Ecological Society of Japan (http://www.esj.ne.jp/esj/), Japanese Institute of Landscape Architecture (https://www.jila-zouen.org/), Crop Science Society of Japan (http://www.cropscience.jp/), and Japanese Society of Environmental Entomology and Zoology (http://kandoukon.org/). The journals published by these societies tend to be associated with particular subdisciplines or taxonomic groups (see AppendixS13 for 40 major Japanese-language peer-reviewed journals relevant to ecology and biodiversity conservation). For example, the Ecological Society of Japan has long promoted plant ecology, with over half of the journal's papers published on plants (Yamamichi & Hiraiwa-Hasegawa, 2012). Such influences might motivate Japanese authors, including local practitioners who are not familiar with English, to publish their research on specific taxonomic groups in the Japanese-language journals of those societies to which they hold membership(s). Second, research questions or hypotheses that require local traditional knowledge may tend to be published in Japanese. In the rice-field meta-analysis, Japanese-language studies provided a particularly high proportion of effect-size estimates for plants (90%; $N = 30$) and amphibians/fish (100%; $N = 13$) (Figure 3). These are species groups that have been most strongly affected by rice-field abandonment in the traditional rural landscapes of Japan (known as *Satoyama* in Japanese)

(Kidera et al., 2018; Koshida & Katayama, 2018). International journals often demand "generality" or "transferability" of study findings, and may judge studies on local topics in Japan as too system-specific. Therefore, studies on plants, amphibians, and fish might be more likely to have been submitted to, and published in, Japanese-language journals. Third, conservation scientists and applied ecologists may choose to publish in Japanese-language journals in order to target a Japanese audience of policymakers and practitioners, who may not read English-language articles. Finally, available resources (time, human, financial) might influence the choice of language for publication. In the leaf life span meta-analysis, most Japanese-language studies measured leaf life span in the field, while most English-language studies employed experimental designs (Figure 4). As measuring leaf life span in the fields is less laborious compared to conducting experiments, this might suggest that studies conducted by more highly funded research groups are more likely to afford English-language proofing, and be published in English-language journals. Those potential processes through which language bias in study characteristics arises will need to be fully investigated in the future.

We found that the differences in effect sizes between the languages remained after controlling for the influence of covariates investigated in the leaf life span and plant forestry meta-analyses. In the leaf life span meta-analysis, between-language differences in effect sizes were apparent even for studies conducted only in Japan, or those from field experiments (Figure6c). This finding might be explained by a general tendency for authors reporting smaller effects in non-English languages, as has been observed in medical sciences (Egger etal.,1997), due to preemption that their results are
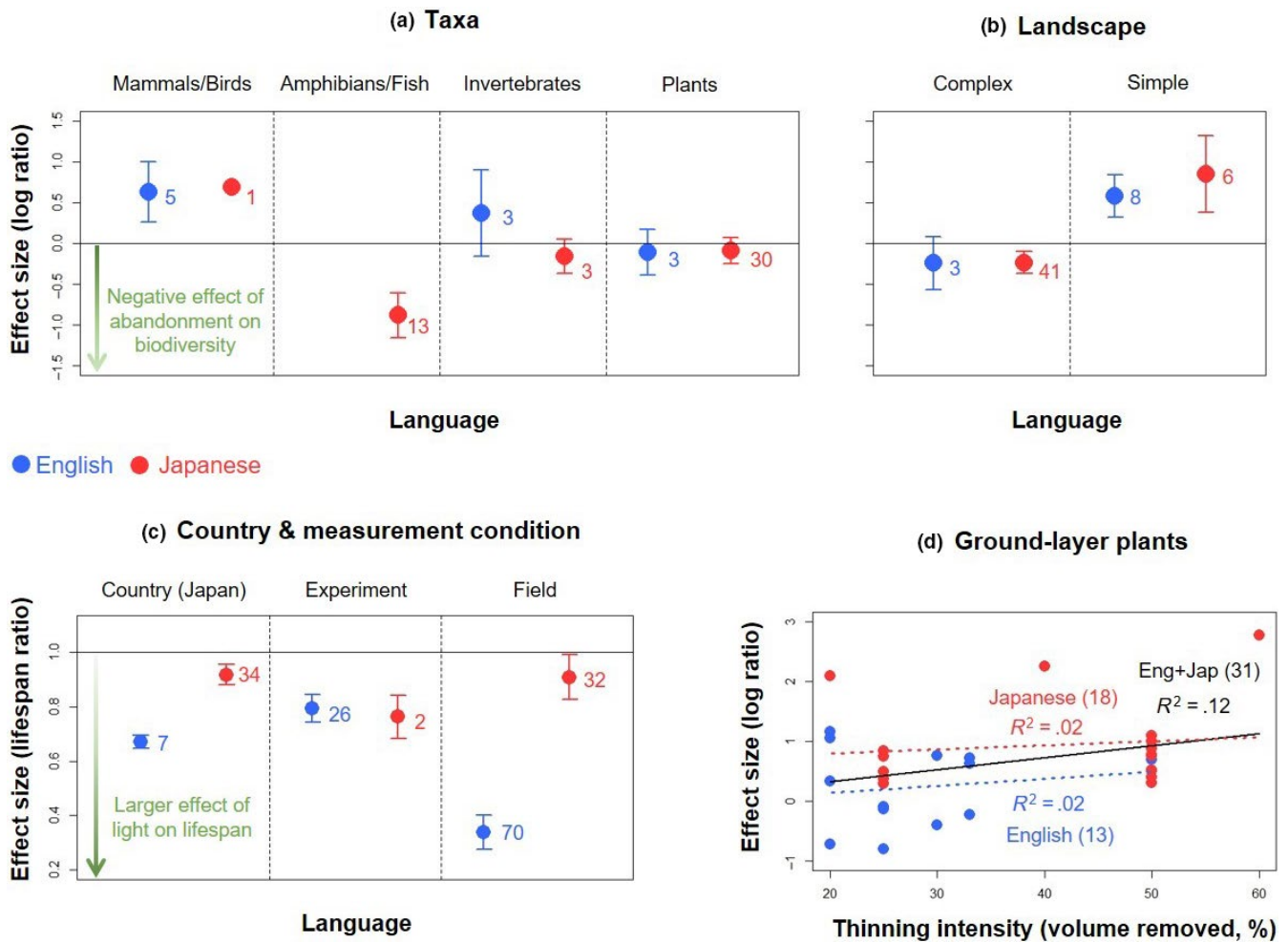
FIGURE 6 Tests for language bias in statistical results after controlling for the differences in study characteristics. Differences in effect sizes between English- (blue) and Japanese-language studies (red) for (a) each taxon and (b) each landscape type in Koshida and Katayama (2018), for (c) studies conducted in Japan, and each measurement condition in Osada et al. (2013), and when (d) the effect of thinning intensity is controlled for in the plant forestry meta-analysis (Spake et al., 2019). The number of effect-size estimates in each language is also shown. The error bars show standard errors. In (d), black line shows the regression line based on all effect-size estimates associated with standard deviations and sample sizes, while red and blue dotted lines show the regression lines based only on effect-size estimates in English- and Japanese-language studies reporting standard deviations and sample sizes, respectively

not strong or interesting enough to be accepted by international English-language journals. Such a tendency is only possible given the availability of local-language journals, where authors can submit their works for publication. Hence, the existence of non-English journals could in effect mitigate publication bias by providing non-English-language speakers with opportunities to report statistically insignificant results, but only if these journals are included in meta-analyses.

Interestingly, we found the opposite effect of language bias in statistical results in the plant forestry meta-analysis; Japanese-language studies had larger mean effect sizes compared to English-language studies (Appendix S12). One possible explanation for this is the issue of pseudoreplication (Davies & Gray, 2015; Spake & Doncaster, 2017), as pseudoreplicated study designs were prevalent in the contributing studies from Japan. While this finding still indicates that ignoring non-English-language studies could bias the estimation of mean effect sizes in ecological meta-analyses, it also

suggests that we should not always assume biases toward a certain direction (i.e., providing larger effects) unless relevant non-English-language studies are identified and incorporated appropriately.

Our study has several important caveats. First, our findings are drawn from only four meta-analyses that met our inclusion criteria. This may partly be due to our criterion for meta-analysis articles to have been published by at least one author affiliated to a Japanese institution. Relaxing this criterion might have increased the sample size. However, even with this restriction, we screened over 1,500 papers and identified 40 potentially relevant meta-analysis articles, of which only three actively searched and included studies published in the two languages. Our small sample size therefore reflects a current common practice of ignoring non-English-language studies in ecological meta-analyses. Second, our findings might depend on how the authors of the meta-analysis articles collected, collated, and appraised the studies for their meta-analyses. However, it was not possible to investigate whether these decisions influenced our

results. Third, within these meta-analyses, we divided studies into two smaller groups according to language, which may have led to increased type II errors of not detecting significant differences (e.g., in Figure 2c,d). Fourth, we explored meta-analyses synthesizing studies published only in English and Japanese, and therefore, biases from other languages remain to be explored. Finally, it is possible that other study characteristics (e.g., geographical difference within Japan), not considered by the original meta-analyses, might have better explained differences among effect sizes between the languages.

Despite being based on a small sample size, our findings have a broad, yet simple, implication for meta-analyses in ecology and conservation science. Future meta-analyses—particularly those conducted at global extents or in regions where English is not widely spoken—should actively search for relevant non-English-language studies and, if appropriate, include them. Searches for non-English-language studies could be implemented by collaboration with native speakers of the relevant non-English languages (Walpole, 2019) or with the aid of emerging technologies (e.g., litsearchr package in R translates search strings into multiple languages: https://elizagrames.github.io/litsearchr/). The flip side is also true; ignoring English-language studies could also introduce language bias, and therefore, even national-level decision-making in non-English-speaking regions requires rigorous assessment of studies in English and in relevant non-English languages. Currently, the use of non-English-language studies in science is not common practice (Baethge, 2013; Neimann Rasmussen & Montgomery, 2018) or even discouraged (Lazarev & Nazarovets, 2018). However, our findings highlight the importance of re-evaluating the role of non-English scientific knowledge in science and a potential risk of ignoring it in meta-analyses. Understanding the generality of our findings across languages, and quantifying any bias incurred from ignoring non-English-language studies, requires further research on other non-English languages (including Spanish, Portuguese, and Chinese; Amano et al., 2016), in which large volumes of scientific literature are published.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTION

Ko Konno: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Visualization (lead); Writing-original draft (lead). **Munemitsu Akasaka:** Conceptualization (supporting); Formal analysis (supporting); Funding acquisition (supporting); Investigation (equal); Methodology (supporting); Supervision (supporting); Writing-review & editing (equal). **Chieko Koshida:** Investigation (equal); Resources (equal); Writing-review & editing (supporting). **Naoki Katayama:** Investigation (equal); Resources (equal); Writing-review & editing (supporting). **Noriyuki Osada:** Investigation (equal); Resources (equal); Writing-review & editing (supporting). **Rebecca Spake:** Investigation (equal); Methodology (supporting); Resources (equal); Writing-review & editing (equal). **Tatsuya Amano:** Conceptualization (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Supervision (lead); Writing-review & editing (equal).

DATA AVAILABILITY STATEMENT

Data are available in Supporting information.

ORCID

Ko Konno https://orcid.org/0000-0002-5319-9178
Munemitsu Akasaka https://orcid.org/0000-0002-0143-0841
Chieko Koshida https://orcid.org/0000-0002-2143-4137
Naoki Katayama https://orcid.org/0000-0002-1136-442X
Noriyuki Osada https://orcid.org/0000-0002-7767-0446
Rebecca Spake https://orcid.org/0000-0003-4671-2225
Tatsuya Amano https://orcid.org/0000-0001-6576-3410

REFERENCES

Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages are still a major barrier to global science. *PLOS Biology*, *14*, e2000933. https://doi.org/10.1371/journal.pbio.2000933

Baethge, C. (2013). The production and recognition of psychiatric original articles published in languages other than English. *BMC Psychiatry*, *13*, 102. https://doi.org/10.1186/1471-244X-13-102

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bayliss, H. R., & Beyer, F. R. (2015). Information retrieval for ecological syntheses. *Research Synthesis Methods*, *6*, 136–148. https://doi.org/10.1002/jrsm.1120

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex: John Wiley & Sons.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*, e1400253. https://doi.org/10.1126/sciadv.1400253

CEE. (2018). *Guidelines and Standards for Evidence Synthesis in Environmental Management VERSION 5.0*. http://www.environmentalevidence.org/information-for-authors

Davies, G. M., & Gray, A. (2015). Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring). *Ecology and Evolution*, *5*, 5295–5304. https://doi.org/10.1002/ece3.1782

Díaz, S., Settele, J., Brondízio, E., Ngo, H. T., Guèze, M., Agard Trinidad, J., … Zayas, C. (2019). IPBES, Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Retrieved from https://www.ipbes.net/news/ipbes-global-assessment-summary-policymakers-pdf.

Egger, M., Zellweger-Zähner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomised controlled trials published in English and German. *Lancet*, *350*, 326–329. https://doi.org/10.1016/S0140-6736(97)02419-7

Grégoire, G., Derderian, F., & Le Lorier, J. (1995). Selecting the language of the publications included in a meta-analysis: Is there a Tower of Babel bias? *Journal of Clinical Epidemiology*, 48, 159–163.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555, 175–182. https://doi.org/10.1038/nature25753

Haddaway, N., Macura, B., Whaley, P., & Pullin, A. (2017). ROSES flow diagram for systematic reviews, Version 1.0. *figshare*. https://doi.org/10.6084/m9.figshare.5897389

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (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, 7. https://doi.org/10.1186/s13750-018-0121-7

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., … Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794. https://doi.org/10.7717/peerj.4794

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]*. London, UK: The Cochrane Collaboration.

Juni, P., Jüni, P., Holenstein, F., Sterne, J., Bartlett, C., & Egger, M. (2002). Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. *International Journal of Epidemiology*, 31, 115–123. https://doi.org/10.1093/ije/31.1.115

Kidera, N., Kadoya, T., Yamano, H., Takamura, N., Ogano, D., Wakabayashi, T., … Hasegawa, M. (2018). Hydrological effects of paddy improvement and abandonment on amphibian populations; long-term trends of the Japanese brown frog, Rana japonica. *Biological Conservation*, 219, 96–104. https://doi.org/10.1016/J.BIOCON.2018.01.007

Koricheva, J., Gurevitch, J., & Mengersen, K. L. (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton, NJ: Princeton University Press.

Koshida, C., & Katayama, N. (2018). Meta-analysis of the effects of rice-field abandonment on biodiversity in Japan. *Conservation Biology*, 32, 1392–1402. https://doi.org/10.1111/cobi.13156

Lazarev, V. S., & Nazarovets, S. A. (2018). Don't dismiss citations to journals not published in English. *Nature*, 556, 174. https://doi.org/10.1038/d41586-018-04169-2

Livoreil, B., Glanville, J., Haddaway, N. R., Bayliss, H., Bethel, A., de Lachapelle, F. F., … Frampton, G. (2017). Systematic searching for environmental evidence using multiple tools and sources. *Environmental Evidence*, 6, 23. https://doi.org/10.1186/s13750-017-0099-6

McAuley, L., Pham, B., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet*, 356, 1228–1231. https://doi.org/10.1016/S0140-6736(00)02786-0

Neimann Rasmussen, L., & Montgomery, P. (2018). The prevalence of and factors associated with inclusion of non-English language studies in Campbell systematic reviews: A survey and meta-epidemiological study. *Systematic Reviews*, 7, 129. https://doi.org/10.1186/s13643-018-0786-6

O'Leary, B. C., Kvist, K., Bayliss, H. R., Derroire, G., Healey, J. R., Hughes, K., … Pullin, A. S. (2016). The reliability of evidence review methodology in environmental science and conservation. *Environmental Science and Policy*, 64, 75–82. https://doi.org/10.1016/j.envsci.2016.06.012

Osada, N., Oikawa, S., Miyata, R., Kamiyama, C., Nagano, S., Shiodera, S., … Ono, K. (2013). Within-species variation in leaf lifespan in relation to environmental factors : General trends and differences across species of different functional types (<Feature 1> Leaf lifespan: Recent advances and future perspectives). *Japanese Journal of Ecology*, 63, 19–36. https://doi.org/10.18960/seitai.63.1_19

Pullin, A. S. (2012). Realising the potential of environmental data: A call for systematic review and evidence synthesis in environmental management. *Environmental Evidence*, 1, 2. https://doi.org/10.1186/2047-2382-1-2

Quinn, G. P., & Keough, M. J. (2002). Experimental design and data analysis for biologists. *Journal of Experimental Marine Biology and Ecology*, 277(2), 197–198. https://doi.org/10.1016/S0022-0981(02)00278-2

R Core Team (2018). *R: A language and environment for statistical computing*. Retrieved from http://www.r-project.org/

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*, 17, 688–690. https://doi.org/10.1093/beheco/ark016

Spake, R., & Doncaster, C. P. (2017). Use of meta-analysis in forest biodiversity research: Key challenges and considerations. *Forest Ecology and Management*, 400, 429–437. https://doi.org/10.1016/J.FORECO.2017.05.059

Spake, R., Yanou, S., Yamaura, Y., Kawamura, K., Kitayama, K., & Doncaster, C. P. (2019). Meta-analysis of management effects on biodiversity in plantation and secondary forests of Japan. *Conservation Science and Practice*, 1, e14. https://doi.org/10.1111/csp2.14

Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology and Evolution*, 19, 305–308. https://doi.org/10.1016/j.tree.2004.03.018

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358, 252–260. https://doi.org/10.1056/NEJMsa065779

Walpole, S. C. (2019). Including papers in languages other than English in systematic reviews: Important, feasible, yet often omitted. *Journal of Clinical Epidemiology*, 111, 127–134. https://doi.org/10.1016/J.JCLINEPI.2019.03.004

WWF. (2018). *Living Planet Report - 2018: Aiming higher*. https://doi.org/10.1080/09528820802312343

Yamamichi, M., & Hiraiwa-Hasegawa, M. (2012). Research object biases in Japanese Journal of Conservation Ecology. *Japanese Journal of Conservation Ecology*, 17, 199–210. https://doi.org/10.18960/hozen.17.2_199

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.