SYSTEMATIC REVIEW

Are small protected habitat patches within boreal production forests effective in conserving species richness, abundance and community composition? A systematic review

Matti Häkkilä[1*] , Anna Johansson[2], Terhi Sandgren[3], Anne Uusitalo[4], Mikko Mönkkönen[1,5], Pasi Puttonen[6] and Sini Savilaakso[2,6]

Abstract

Background: In boreal zone forest management is changing and degrading forest habitats, which has caused declines in biodiversity. To mitigate these harmful effects in production forests, small-scale habitats with high biodiversity values have been protected within them. These habitats include woodland key habitats, and other small habitat patches protected by voluntary conservation actions. In this systematic review we synthesize the evidence on the value of small protected habitat patches (SPHP) within production forest landscapes for biodiversity. Review question: Are small protected habitat patches within boreal production forests effective in conserving species richness, abundance, and community composition?

Methods: Both peer-reviewed and grey literature were searched from bibliographical databases, organizational websites and internet search engines in English, Finnish, Swedish and Russian. Articles were screened at two stages (title/abstract and full text) and the validity of the included studies were assessed. Screening and validity assessment were based on predetermined criteria. After data extraction, narrative and quantitative syntheses were conducted. Influences of effect modifiers were tested, and sensitivity analyses were conducted.

Review findings: During the searches 19,458 articles were found. After duplicate removal and title/abstract screening 336 articles remained. During full text screening 41 articles were included and 35 of them (174 studies) were included in narrative synthesis. 28 articles with 127 studies had suitable data for meta-analysis. SPHPs had significantly higher species richness compared to production forests. When compared to natural forests, there was no significant difference. Forest management in areas surrounding SPHPs did not have impact on species richness of these patches. Individual abundance was significantly higher in SPHPs compared to natural or production forests. There was significantly more dead wood in SPHPs compared to production forests, but when compared to natural forests there was no significant difference. Community composition was different between SPHPs and both production and natural forests.

*Correspondence: matti.i.hakkila@jyu.fi
[1] Department of Biological and Environmental Science, University
of Jyväskylä, P.O.Box 35, 40014 Jyväskylä, Finland
Full list of author information is available at the end of the article

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 2 of 20

Conclusions: The findings of this review show that small protected patches within production forests are important part of biodiversity conservation. They cannot substitute larger protected areas but supplement the protected area network. However, there were gaps both in geographical distribution of the studies as well as in the selection of target species of the studies. Therefore, generalization of the results must be done carefully.

Keywords: Forest harvesting, Logging, Impact, Woodland key habitats, Voluntary conservation, Species richness, Species diversity, Abundance, Assemblage, Dead wood

Background

Protecting areas from human usage has traditionally been the main method for biodiversity conservation. However, setting aside areas in productive forest land is expensive and only a relatively small proportion of forests is protected for biodiversity in Northern Europe [1] Globally, the network of protected areas is considered inadequate to maintain species assemblages [2–5]. Protected areas are often situated in less productive areas with low biodiversity [6, 7], and establishing new protected areas is not possible in the magnitude needed for maintaining biodiversity [5]. Therefore, conserving biodiversity in production forests has become increasingly important.

Practices to maintain and increase biodiversity in production forests include green tree retention, prescribed burning, leaving dead wood in forests and creating habitat corridors and buffer strips [7, 8]. In addition, preservation of small patches of certain habitats, e.g. woodland key habitats (WKHs), was introduced as a method in biodiversity conservation in early 1990′s [9]. Woodland key habitats are a common concept in Northern Europe first introduced in Sweden [10, 11]. Nowadays the concept is used also in Finland, Norway, the Baltic countries, and Russia (mostly in the northern regions of the European part of the country and Siberia [12]) There are differences in the definitions and legal status of the WKHs and their delineation vary between countries [9–20]. For example, in Sweden, WKHs are defined by certain structural properties or species present in them and the protection of WKHs is mostly voluntary while in Finland WKHs are protected by the Forest Act [10, 14].

In addition to WKHs, other new kinds of conservation policies have been developed. Finnish *Forest Biodiversity Program for Southern Finland* (METSO) [21], Norwegian *Frivillig vern* [22] and Swedish *Komet programmet* [23], are all based on the forest owners´ own initiative to protect their forests. Forest owners offer their forest to be protected by the program, and if the forest has sufficient ecological values, e.g. high volume of dead wood or quantity of large deciduous trees, the owners will get a compensation payment.

In the USA the concept of conservation easements is the main way of voluntary forest conservation. Conservation easements mean that landowners give up their right to develop the forest and they get monetary or taxation-based compensation from the government or a conservation group (land trust) [24]. In Canada and Russia, almost all forests are publicly owned, and therefore comparable voluntary conservation systems do not exist. In the Baltic countries the private forest ownership has been re-established after regaining their independence in 1991 [25]. Environmental values of forests are emphasized both by national forest programs and private forest owners, but most conservation programs are still determined by state authorities [25, 26].

Forest management certification systems may also include provisions for conserving certain habitats beyond legal requirements. Both Forest Stewardship Council (FSC) and Program for the Endorsement of Forest Certification (PEFC) systems are widely used in the boreal zone [27, 28]. To get certified, forest owners must commit themselves to responsible forest management. National certification standards vary between countries, but they often include preservation of small habitat patches of high conservation value [29–31].

But is conserving small habitat patches within production forests an effective way to maintain biodiversity? There has been critique of the small size and scattered distribution of WKHs [32]. It has been suggested that isolated woodland key habitats suffer from extinction debt [33] and that small habitat patches may not be able to maintain species diversity over time [34]. In previous systematic review, WKHs were found to be high in species richness and in the number of red-listed species but their ability to maintain diversity in managed landscape could not be addressed [35]. As the evidence base has not been reviewed recently, it is time for an update. Here we present a systematic review on the importance of small protected habitat patches (SPHPs hereafter) within production forests for maintaining biodiversity. The topic arose from the interest of Finnish forest industry on the effectiveness of conservation actions within production forests to produce beneficial biodiversity outcomes. A stakeholder workshop was held in November 2018 to further discuss the topic. There was a balanced representation of different stakeholder interest groups from government agencies and academia to non-governmental organizations and the private sector. In the workshop

Fig. 1 Simplified theory of change on the contribution of different actions to maintain or increase biodiversity in production forests to counterbalance the effects of forest harvesting
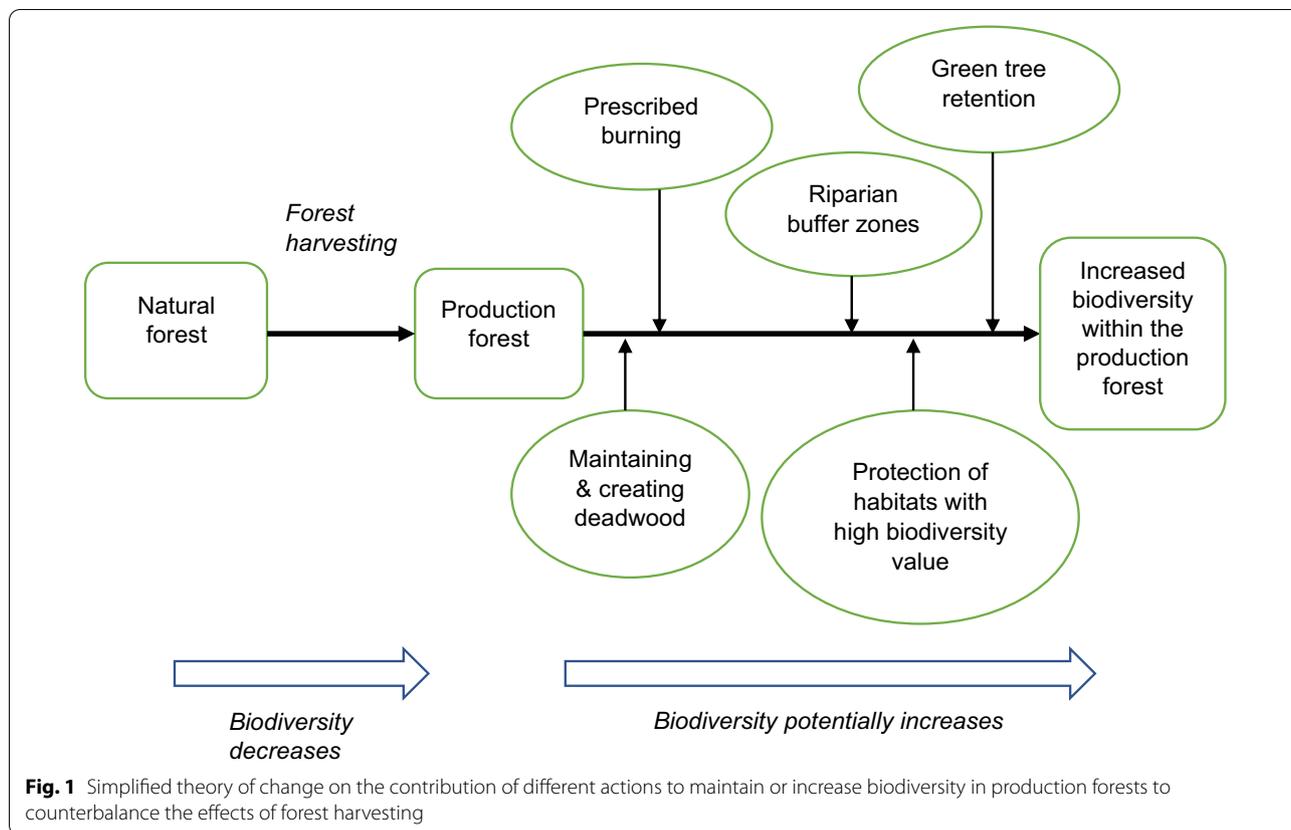
review questions initial theory of change (Fig. 1) and PICO-based search terms were defined and factors potentially creating heterogeneity discussed. Further details of the stakeholder engagement are provided in the protocol [36].

Objective of the review

The objective of this article is to systematically review and synthesize the impacts of conservation of small habitat patches within production forests on biodiversity. Fundamentally, biodiversity is a complex measure at genetics, species and ecosystem level, but here we concentrate on species level diversity, yet still calling it biodiversity. Legally designated woodland key habitats as well as any small-scale voluntary conservation areas are included. The review focuses on terrestrial biodiversity in boreal forests. We follow the definition of boreal zone by Keenan et al. [37] but will also include Baltic countries. The geographical scope was determined at the stakeholder meeting.

Review question:

Are small protected habitat patches within boreal production forests effective in conserving species richness, abundance, and community composition?

The question components are:

P: Boreal forests (Canada, USA, Norway, Sweden, Finland, Russia, Estonia, Latvia, Lithuania).

I: Small habitat patches set aside for conservation within production forests (SPHPs).

C: Managed forests, natural forests.

O: Species diversity, abundance, assemblage, dead wood volume and diversity.

Methods

This systematic review follows the guidelines of Collaboration for Environmental Evidence and complies with the ROSES reporting standards. The ROSES form is included as an Additional file 1. All the deviations from the a priori systematic review protocol [36] are reported and explained below.

Deviations from the protocol
Search string

After the publication of our systematic review protocol [36] we noticed some mistakes in the search string and few changes were made. The retention term was

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 4 of 20

mistakenly included in the search string when the protocol was prepared for publication after the peer review.

Internet searches

In internet searches, one addition to the search method was made compared to the protocol [36], as no citations or patents were considered in Google Scholar.

Validity assessment

Because of limited time resources, all studies were assessed by one person and their decision checked by another person resulting in a small deviation of the protocol [36] as double-blind assessment was not used.

Potential effect modifiers and sources of heterogeneity

Two effect modifiers were added to the list in the protocol [36]: altitude and differences in the forest management surrounding SPHP, and one was removed: size and disposition of retention trees because this information was merged with another effect modifier: differences in forest management (comparator). These effect modifiers were added during the data extraction when differences concerning these potential effect modifiers were noted between included studies.

Searching for articles

Articles were searched from 1990 onwards because small scale habitat protection within production forests was integrated in production forestry in the 1990s.

Search terms

Based on the discussions at the stakeholder meeting, a search string was formulated. The performance of the search string was tested in the Web of Science and Scopus using a test list of 20 articles collected from previous reviews and from experts. The search string used for testing in the protocol [36] mistakenly included term TS = (retention*) and, also the test list in the protocol mistakenly included three articles dealing with retention as originally a broader scope was proposed for the review. These mistakes were caused by human error when modifying the manuscript after peer review. Since retention was not included in the final manuscript, the three articles about retention were removed from the test list and the performance of the final (corrected) search string was re-tested. The corrected test list is in Additional file 2 and the results from testing of the search string in Additional file 3.

The final search string (Web of Science format) was:

#1 TS = ((Boreal NEAR/5 (forest* OR zone OR tree*)) OR taiga OR spruce* OR picea OR pine* OR pinus OR birch* OR aspen* OR populus)

#2 TS = (Finland OR Finnish OR Swed* OR Norw* OR Russia* OR Estonia* OR Latvia* OR Lithuania* OR Fennoscan* OR Scandin* OR Baltic OR "North* Europ*" OR Canad* OR "North* Ameri*" OR Siber* OR Alaska OR "United States" OR USA) and TS = (forest* OR tree*)

#3 = #1 OR #2

#4 TS = ("key habitat*") OR TS = ("forest act habitat*") OR TS = (reserve* NEAR/5 (forest* OR OR privat* OR area* OR patch* OR habitat*)) OR TS = ("private* protected area*") OR TS = (voluntar* NEAR/5 (conservation* or set-aside*)) OR TS = (METSO NEAR/5 program*) OR TS = (Komet NEAR/5 program*) OR TS = (conservation NEAR/5 easement*) OR TS = (connectiv*)

#5 = #3 AND #4

The search string was translated to Russian. A simplified version of the search string containing only key terms was used when the search interface did not have the capacity to handle the whole search string. The simplified search strings were translated to Swedish and Finnish when needed. Boolean operators were used where possible. Final search strings used are reported in Additional files 4, 5. To detect articles that were published before the data synthesis was started, a search alert was set in two bibliographic databases (Russian Science Citation Index on the Web of Science (RSCI) and Scopus) and Web of Science Core Collection (WoS)), which consists of six databases: Science Citation Index Expanded (1945-present), Social Sciences Citation Index (1956-present), Arts & Humanities Citation Index (1975-present), Conference Proceedings Citation Index-Science (1990-present), Conference Proceedings Citation Index-Social Science & Humanities (1990-present), Emerging Sources Citation Index (2015-present) (hereafter treated as one for simplicity).

Languages

This systematic review includes studies published in English, Finnish, Swedish and Russian. The language selection is based on the geographical scope of the systematic review and is limited by the language skills of the review team. Organizational websites and bibliographic databases were searched in English, except of Finnish, Swedish and Russian ones that were searched in the primary language the website was published. In addition, if the publications section / library catalogue included studies published in other of the review languages (e.g. main website language is Swedish but there are also unique publications in English), the search was conducted in those languages as well. The searches in search engines were conducted in all four languages.

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 5 of 20

Bibliographic searches

The searches in RSCI, Scopus and WoS were conducted using institutional subscriptions of the University of Helsinki and the full search strings in English and Russian were used. For the rest of the databases, no subscriptions were needed, and simplified search strings adjusted for the functionality of the search field was used. For the full search details see Additional file 5.

Search engines

Searches with internet search engines were conducted in 'private' mode to prevent the influence of previous browsing history and location on search results. The searches were conducted in English, Finnish, Russian and Swedish with simplified search strings (Additional file 4). Results were organized by relevance. After the first 50 hits, results were checked until relevant articles were no longer retrieved as advised in Livoreil et al. [38]. To safeguard against finishing the search too early, a hundred irrelevant hits were allowed. If no relevant hits appeared after a hundred irrelevant hits, the search was terminated. The maximum amount of hits shown in Google and Google Scholar is 1000, so this was also the maximum amount of screened hits per search. For the full search details see Additional file 5. The following search engines were used:

> Google Scholar (https://scholar.google.com/)
> Google (https://www.google.com/)

Organizational websites

Besides known sources of potential literature, internet was searched extensively to find additional sources. Organizational websites dealing with related issues were chosen and scoping for possibly relevant material was conducted before decision on which organizational websites will be searched was made. Searches were conducted in the publications section of the website, if available. If not, the search was conducted using the "search" function of the front page. In cases where the search option was missing or it was obviously not working correctly, the websites were searched manually. Also, all Russian webpages were searched manually. The search strings/terms were adjusted to each website according to its functionality. The final list of organizational websites and the full search details in each website are given in the Additional file 5.

Supplementary searches

To supplement the search, citation chasing was undertaken in the relevant review articles included at title/abstract stage but excluded at full text stage (the reviews, where citation chasing was conducted are identified in Additional file 6). Also, a call for unpublished data was published on the website of the Evidence-Based Forestry in Finland project (http://npmetsa.fi/en/frontpage/) and in ResearchGate and sent directly to stakeholder organizations that may have unpublished data on the topic.

Article screening and study eligibility criteria
Screening process

Articles were screened in two stages by three people: first by title and abstract (MH, AJ, SS) and then by full text (MH, AJ). At the first stage a random set of 100 articles was independently screened by all three persons. If their inclusion decisions did not vary more than 5%, the screening process could commence. If the screening decisions differed more than 5%, a second set of 100 articles were independently screened by all three persons. This procedure continued until at least 95% consensus was achieved. All discrepancies in screening decisions after one set of 100 articles were discussed to facilitate consistency in the screening process. If a reviewer was unsure whether to include or exclude an article on the title/abstract stage, the article was included.

At the full text stage, the articles were independently screened by three people (MH, AJ, SS) All the articles that a reviewer was unsure about were discussed with other research group members. Articles in Russian were screened by only one person (AJ). As they were screened after testing consistency of reviewers' decisions, the Russian speaker talked through the decision process of the articles she was unsure about. In most cases it was clear whether the article fulfilled the inclusion criteria or not.

Authors of this review who had also authored papers considered in this review did not take part on the article screening.

Full texts of articles included at the title/abstract stage were searched in the internet, Helsinki University Library, Jyväskylä University Library and by interlibrary loans. In the case an article was not found through these searches, it was classified as unretrievable. The inclusion of articles published by the authors of the review at the screening and critical appraisal stage was jointly determined by the other review group members in accordance with the eligibility and appraisal criteria.

Search records from WoS and Scopus were exported into the reference management software EndNote. The files were merged, and duplicates were removed. To conduct screening, articles were exported into Colandr, a machine-learning assisted application for conducting systematic reviews [39]. Search records from other databases were collated and the articles were screened manually.

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 6 of 20

Table 1 The eligibility criteria for article screening for the study question

Question elements	Eligibility criteria
Populations	*Included* Boreal forests in Canada, Estonia, Finland, Latvia, Lithuania, Norway, Russia, Sweden, and USA (dominant tree species in the study area must be spruce or pine)
Intervention	*Included* Conservation of woodland key habitats Conservation of the small protected forest patches, e.g. voluntarily conserved small areas (METSO and Komet programs) and certification-based protection *Excluded* Conservation of large protected areas, i.e. national parks, nature reserves, wilderness areas
Comparators	*Included* Managed forests (production forests of all type, age and succession stage) Natural forests (including national parks and other large-scale protected areas and, also, non-protected natural forests) Also, SPHPs in natural and production forests can be compared to each other *Excluded* Small protected habitat patches different from intervention (for example WKHs and voluntary set-asides are not compared) Different small protected habitat patches within the intervention (for example stream-side WKHs and herb-rich WKHs not compared) Managed forest known to have low/high conservation values (for example areas not accepted in METSO program, forest areas receiving environmental subsidy) Non-forest lands Urban parks Wooded fields e.g. Christmas tree plantations
Outcomes	*Included* Species diversity Species richness Species abundance Assemblage Volume of deadwood Diversity of deadwood *Excluded* All data on non-terrestrial species Density of ant mounds / predator bird nests
Study design	*Included* Control-intervention field studies of small habitat patches Control-intervention studies of small habitat patches where no field study was conducted but the data analysed are originally from a field study *Excluded* Simulation/modelling studies (even partly simulated/modelled) Habitat selection studies Studies first mapping/finding species and afterwards checking whether the place of occurrence was located in SPHP or not
Language	*Included* English Finnish Swedish Russian

Eligibility criteria

To be included in this review, screened articles had to pass eligibility criteria based on PECO-components, study design and language (Table 1). The criteria were specified more in detail after publication of the research protocol [36].

Study validity assessment

All studies included in the full text stage were critically appraised and categorized. Critical appraisal was based on study design, sampling, accounting for potential effect modifiers and data analysis methods. These were considered by the authors of this review to be key variables related to generalisability and reliability of study findings. The criteria related to risk of bias and confounding factors (internal validity) and generalisability of the results (external validity). They include, for example, appropriate and representative replication, suitable outcome measuring methods, possibility of researchers to control baseline differences, measures taken to address potential baseline differences (e.g. accounting for spatial heterogeneity), and information about potential confounding factors) (Table 2). Following the critical appraisal, studies were categorised as having 'low',

Table 2 Critical appraisal criteria to assess studies in the full text stage

Factor	Low risk	Medium risk	High risk
Study design	Experimental studies (includes also quasi-experimental studies)	Observational studies	Case studies (descriptive studies)
Sampling	Large sample size relative to outcome measure and species in question (high confidence that replication is appropriate and representative) Sampling method suitable for the population of interest[a] Randomisation accounting for spatial heterogeneity Control and intervention areas matched based on their ecological characteristics	Small to medium sample size relative to outcome measure and species in question (medium level of confidence that replication is appropriate and representative). Sampling method suitable for the population of interest[a] Control and intervention areas comparable based on their ecological characteristics	Sampling method not suitable for collecting data on the population of interest[a]
Accounting for heterogeneity and potential effect modifiers	Potential biologically important effect modifiers that could influence the study findings identified, and data collected on them. The context in which the study took place clear	Potential biologically important effect modifiers that could influence the study findings identified, and considered in relation to the results. The context in which the study took place clear even when there was no direct data	Effect modifiers not identified or considered. The context of the study is unclear
Data analysis methods	Methods appropriate[b]	Methods appropriate[b]	Methods not appropriate[b]

[a] Suitable sampling method refers to the use of methods that are known to work for the population in question based on published studies, e.g. flying insects are sampled by trapping or fogging, not by cutting branches

[b] Appropriate methods refer to the use of statistical methods that consider data characteristics such as sample size and distribution. For example, non-parametric statistical tests are used for data that does not follow normal distribution

'medium', or 'high' risk of bias. If information was inadequate to make an assessment, the study was labelled as 'unclear'. Studies in 'low' and 'medium' categories were considered to have sufficient quality to provide reliable evidence base for quantitative synthesis and hence, the categorisation was used mainly in the narrative synthesis.

The studies were categorized based on their lowest performance, i.e. if any of the factors falls in the category 'high risk', the study will be categorized as high risk of bias. No studies were excluded based on validity assessment, but during the meta-analysis the studies were weighted according to their category (see data synthesis and presentation section for testing robustness of results with sensitivity analysis). The validity assessment was conducted at the same time with data extraction by two reviewers (MH, AJ). Any inconsistencies or uncertainties were discussed with other research group members.

Data coding and extraction strategy

When there were multiple articles from one study site (i.e. linked articles), they were appraised as a group to avoid inclusion of duplicate data following Frampton et al. [40]. When there were results from several taxa in one article, these were treated as separate studies in data extraction. Data from included studies were extracted and saved in an Excel spreadsheet and are available as Additional file 7. To ensure consistency during data extraction, MH and AJ extracted data from five articles together before beginning the extraction independently.

Basic publication details and information about intervention, comparator, outcome, study subject (species) and study design as well as brief description of the sampling method were extracted. For quantitative analysis, outcome means, standard deviations (SDs), standard errors (SEs), and sample sizes were extracted from text, tables, and graphs. Image analysis tool WebPlotDigitizer [41] was used where necessary. In cases where information on means, SDs or SEs were not available, data on test statistics that could be used in conversions were collected. For analysis of assemblage data, total number of species (intervention + comparator) and number of shared species between intervention and comparator or other description of the community similarity were extracted. Furthermore, data on effect modifiers and potential sources of heterogeneity were extracted whenever available to enable statistical analysis of the relationships between outcomes and sources of heterogeneity. The management of the intervention surroundings was determined as intensive (clear cut, sapling stand after clear cut), moderate (middle-aged or mature even-aged production forest stand) or light (uneven-aged production forest or near natural stand).

In many articles there were data on all species and rare species, i.e. red-listed / indicator / other rare species. In those cases, data was extracted only on rare species as the purpose of small protected habitat patches is to protect especially rare, endangered species. By extracting data specifically on those species, conclusions can be drawn

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 8 of 20

how well small protected habitat patches fulfil their purpose. No data on single species were extracted, but data on separate species groups (bryophytes, lichens, vascular plants etc.) were extracted when available. Two kinds of dead-wood data were extracted: volume and diversity. Whenever given, the total volume was extracted.

If an article had studies from various locations, the data on results combining these locations were extracted. If combined data were not available, data on all locations were extracted as different studies. However, the locations had to be far enough from each other based on the dispersal ability of the taxa in question to ensure results were independent. If there were studies from multiple years, the data on results combining these years were extracted. If combined data was not available, only data from the latest year were extracted (to avoid non-independence of the data).

Data were extracted by two persons (MH, AJ). Thus, to ensure consistency, a set of five studies were first coded together. Any uncertainties with data extraction were discussed among group members. To retrieve missing information or data, authors of the articles were contacted via email.

Potential effects modifiers and sources of heterogeneity
To understand better possible variation in the effects of the studies, possible effect modifiers were extracted from the studies. The list below was compiled based on the authors' experience and consultation at the stakeholder meeting and it was peer reviewed and published in the protocol of this review [36].

- Geographic location
- Study year
- Climatic conditions
- Forest type
- Soil type
- Altitude
- Differences in forest management (comparator)
- Differences in the forest management surrounding SPHP
- Tree species composition
- Size of trees
- The category and size of SPHP
- Certification (certified or not, certification system)
- The owner of the study site(s)

Data synthesis and presentation
A narrative synthesis of data from all the individual studies included was produced. All the information described

in the narrative synthesis can be found in Additional files 6 and 7.

In addition to the narrative synthesis, a quantitative synthesis, i.e. meta-analysis, was conducted to assess the effects of protection of SPHPs on biodiversity. Studies that provided quantitative data on outcome measures were included in the meta-analysis. Studies that had incomplete or missing information that could not be retrieved were excluded from the meta-analysis.

Meta-analysis
We used the standardised mean difference (Hedges' d) as a measure of the effect size for species richness, abundance, and deadwood volume:

$$d = \frac{(X_i - X_c)}{s}J$$

where X_i and X_c were the means of the intervention (SPHP) and control groups (production and natural forest), s is the pooled standard deviation and J is a correction term mitigating for small sample size bias. The pooled standard deviation was calculated as

$$s = \sqrt{\frac{(n_i - 1)SD_i^2 + (n_c - 1)SD_c^2}{n_i + n_c - 2}}$$

where n_i and n_c are the sample sizes of the intervention and control groups and SD is the standard deviation. Correction term J [42] was calculated as

$$J = 1 - \frac{3}{4(n_i + n_c - 2) - 1}$$

Variance for Hedges' d was calculated as

$$var = \frac{n_c + n_i}{n_c n_i} + \frac{d^2}{2(n_c + n_i)}$$

If there were non-independent cases within a study (i.e. SPHP was compared to both production and natural forest), we calculated a corrected overall sample size, $N_{corrected}$, to avoid double counting the intervention sample size following the method described in Gleser and Olkin [43]:

$$N_{corrected} = n_i + \sum_1^i n_c$$

where n_i is the sample size of the intervention and n_c are sample sizes for the controls.

$N_{corrected}$ was then used to calculate corrected S_{pooled}, J, *Hedges' d* and its variance for studies with non-independent cases:

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 9 of 20

$$S_{pooled.corrected} = \sqrt{\frac{(n_i - 1)SD_i^2 + (n_c - 1)SD_c^2}{N_{corrected} - 2}}$$

$$J_{corrected} = \left[1 - \frac{3}{4(N_{corrected} - 2) - 1}\right]$$

$$var_{corrected} = \frac{1}{n_c} + \frac{1}{n_i} + \frac{d^2}{2(N_{corrected})}$$

If information on means, SDs or SEs was missing, we used available test statistics to either calculate SDs or to convert available test statistics to Hedges' d. In the deadwood data set, we used imputation to calculate missing SDs for five studies following Lajeunesse [44]:

$$\widetilde{SD} = \underline{X_j}\left(\frac{\sum_i^K SD_i}{\sum_i^K \underline{X_i}}\right)$$

where X_j, is the observed mean of the study with missing information, and K is the number of *jth* studies with complete information (*i*). The imputation was done separately for intervention and comparator arms of the study and per comparator forest type to minimise biases.

Meta-analyses were conducted in R [45] using *the rma.mv* function in the *metafor* package [46]. The *rma.mv* function allows fitting of a multi-level model with a nested structure with or without moderators. In our models, studies were nested within articles as there were articles, which reported multiple outcomes. We used profile likelihood plots to check that both variance components were identifiable by the model. For deadwood, study level variance was estimated to be zero in the model, which means that the model corresponds to a model with one level only. We used inverse-variance weights in the model, i.e. weights were equal to $w^i = (v_i + \tau^2)^{-1}$. Cochran's Q-test [47] was used to test whether the variability in the observed effect sizes is larger than it would be expected based on sampling variability alone. A significant test indicates that the true effects are heterogeneous. When moderators were included in the model, the Q-test tests for residual heterogeneity, i.e., remaining variability in effect sizes after variability due to moderators is accounted for.

For each outcome variable (species richness, abundance, and deadwood) we first calculated the overall mean effect size. If significant heterogeneity in the effect sizes was found, we explored the influence of effect modifiers thought to be biologically significant in separate models. Following models were performed for species richness: 1. comparator, taxa and their interaction, 2. influence of forest owner, 3. intensity of management of the area surrounding SPHPs, and 4. age

of the comparator forest. For the models 2–4 subsets of data were used as not all the studies reported same information and combining them in same model would have resulted in a significant loss of power. For abundance, we tested the influence of taxa and comparator only as the number of studies per comparator forest category was limited (n = 15). For deadwood, we tested only whether age of the comparator forest influenced effect sizes as the dataset was very small (n = 7).

In addition to the models above, we explored publication bias in the datasets by producing funnel plots and conducting trim and fill-tests. If outliers were found, we tested their influence on the results. We also tested whether effect sizes were influenced by study characteristics: the type of publication (MSc thesis, journal article, report), country or year when data was collected. We conducted sensitivity analyses by excluding studies that had corrected sample sizes, imputed SDs or were studies with 'high risk' of bias. The script used to run models in R is provided in Additional file 8 and the data used in these models is provided in Additional file 7.

Analysis of the assemblage data

A variety of different methods used to examine changes in species composition makes it difficult to quantitatively assess the effects of habitat modification on species composition. Hence, to have a standardized measure to assess changes in species composition, we used a simple averaging method following Nichols et al. [48]. For each study, we first extracted or calculated the number of shared species between SPHPs and comparator forests. Then we divided the number of shared species by the total number of species recorded in SPHPs. To know the overall response, we calculated the mean of all the studies and its 95% confidence intervals. The overall response was considered significant when the confidence interval did not include one.

Review findings
Review descriptive statistics
Search and screening of articles
Searches in Scopus and WoS were conducted from December 2nd to 3rd, 2019 resulting in 8,303 hits. Other database searches were conducted December 12th to 17th 2019 producing 4 787 hits. Search engine searches resulted 4250 hits between February 4th to 17th 2020 and searches in organisational websites (not Russian) resulted in 1914 hits during February 7th to 12th 2020. The search alerts were on from December 2nd, 2019 to March 19th, 2020 producing 201 hits. The manual searches in Russian databases produced 3 articles. Altogether the searches resulted in 19,458 hits. Duplicates were first removed from the records imported to EndNote (hits from Scopus

and WoS). Because articles from other sources could not be imported to EndNote, they were screened manually, and duplicates were removed during title/abstract screening. After duplicate removal and screening by title and abstract, 336 articles remained (Fig. 2). No articles were obtained through call for unpublished data. One set of 100 randomly chosen articles was needed to achieve 95% consistency in the screeners' (MH, AJ, SS) decisions at the title and abstract stage.

Thirteen full texts were not found which means that a total of 323 full texts were read. After full text screening 38 articles were included. In addition, two articles were added during citation chasing, and one article was found by chance encounter outside the predetermined sources increasing the total number of articles included by full text to 41 (Fig. 1). The most common reason for exclusion was that there was no eligible intervention (no small-scale protected habitat patches or these patches were combined with other forest types). Also, the lack of suitable comparator was a common reason for exclusion (for example articles, where all the studied areas were SPHPs). The list of all articles included and excluded by full text with the reasons for exclusion is given in Additional file 6 and a description of the search of each database and the articles included in each screening stage is given in Additional file 5.

Of the 41 articles included at the full text stage, 20 belonged to some group (articles from the same study site formed a group). There were eight separate groups with 2–4 articles in each. Despite the substantial number of articles belonging to a group, a total of 35 articles had independent data. Therefore, the number of articles and studies included in narrative synthesis was 35 and 174, respectively. Some studies included in narrative synthesis did not have suitable data for further data synthesis and eventually 28 articles with 127 studies were included for quantitative analysis (Fig. 1).

Of the articles included at full text stage, three were authored by an author of this review (MM). Two of these had independent data and were also included in the quantitative analysis. The screening and critical appraisal for these articles were conducted by MH, SS and AJ (Fig. 2).

Sources of the articles included in narrative synthesis
The majority of the included articles (22 of the 35 articles, 63%) were found through searches in Scopus and WoS. Three articles (9%) were found in other bibliographic databases, six (17%) in Google Scholar and two (6%) in organisational websites. Citation chasing and other sources both resulted in one article.

There were three types of articles: journal articles, master's theses, and reports. 25 articles were journal articles, six were master's theses and four were reports (25 peer reviewed and 10 grey literature articles). In addition to WoS and Scopus, journal articles were found in Google Scholar (Finnish search), Doria (Finnish search) and from other sources, one from each source.

Most of the articles were written in English (27 out of 35). Four were written in Finnish and four in Swedish. All articles written in Russian were excluded during the full text screening at the latest. The search was limited to studies published after year 1990, however, no articles written in the 1990's were included. 22 articles were published in years 2000–2009 and 13 in years 2010–2019.

Narrative synthesis including validity assessment
Locations
Most of the studies were conducted in Sweden, followed by Finland, Norway, and Latvia. No studies from other eligible countries were included (Fig. 3).

Interventions
Almost all the studies considered were WKHs. There were only five studies from two articles where the areas were not WKHs but protected by METSO program [50] or by certification [51]. In most cases, it was not written what kind of WKHs were studied but stream-side forests [52–56] and herb-rich forests [57] were mentioned.

Comparators
In 113 studies (65%) the comparator was production forest, whereas in 61 (35%) it was natural forest. Production forest comparators were mostly described as 'ordinary managed forest' which usually means even-aged forest where thinning has been done and which consists of only few tree species. Age of the production forest was not mentioned in majority of the articles but where it was, it varied between 50 [53] and 120 years [50]. In one article [58] production forest comparator consisted of a mix of clear cuts and older forests. The natural forest comparator mostly referred to protected natural reserves and national parks. Six articles (with 42 studies) had both production forest and natural forest comparators [51, 59–63].

Outcomes
Species diversity (i.e. richness) was the outcome in 61 studies (35.1% of the studies). The number of studies for other outcomes were: assemblage 44 (25.3%), individual abundance 39 (22.4%), dead wood volume 27 (15.5%) and dead wood diversity 3 (1.7%). Studied subjects could be divided into eight groups (Fig. 4). The most studied

Fig. 2 Flow diagram adapted from ROSES [49] showing literature sources and inclusion/exclusion process. Note that duplicate removal was conducted at two stages. Duplicates were first removed from the Scopus and WoS databases after the records were exported to EndNote. Because records from other searches could not be exported to EndNote, duplicate removal was conducted at title/abstract stage

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 12 of 20

Fig. 3 Number of studies included in narrative synthesis by country

Fig. 4 Number of studies by study subject

subject was fungi (mostly polypores) and vascular plants followed by dead wood and bryophytes. 100 studies (57%) focused on rare (red-listed) species, indicator/signal species or species associated with dead wood.

Potential effects modifiers and sources of heterogeneity

Information about potential effects modifiers and sources of heterogeneity was mostly insufficient (Additional file 7). Geographic location, study year and forest type were reported in most studies. Climatic conditions, soil type, altitude, category and area of SPHP, age of the forest, certification status and owner of the forest were reported occasionally.

Study validity assessment

Almost all studies (168 of 174) were categorised as medium risk of bias during the critical appraisal and the rest were assessed as having a high risk of bias because

no effect modifiers were identified [64, 65] or because the intervention and comparator data were collected with different methods [66]. The studies with high risk of bias were not excluded from the quantitative analysis but their influence on results was tested by conducting a sensitivity analysis. The reason for absence of studies with low risk of bias was the lack of experimental studies. The critical appraisals for all studies are reported in Additional file 7.

Data synthesis

Description of studies included in the quantitative synthesis

127 studies (from 28 articles) had suitable data for quantitative analysis. These evidences represent 73% of all the studies included in the narrative synthesis. The outcome was species richness in 45, abundance in 30, assemblage in 26 and dead wood volume in 26 studies. There were not enough data to conduct meta-analysis on dead wood diversity. All the studies where intervention was not WKH (but METSO habitat or protected by certification) had dead wood volume as outcome.

75 studies compared SPHP to production forest and 52 studies were comparing SPHP to natural forest. 63 of the studies were conducted in Sweden, 49 in Finland, 15 in Norway and none in Latvia. Vascular plants and fungi were the most studied taxa (31 and 25 studies, respectively). 72 studies were published in journal articles, 49 in master's theses and 6 in reports. Three studies were assessed as having a high and the rest as having a medium risk of bias. In all three high risk studies the outcome was dead wood volume.

Species richness

The overall mean effect size was significantly different from zero ($d = 0.812$, $p < 0.0001$, 95% CI 0.407, 1.217) meaning that SPHPs had significantly more species than comparator forests. There was significant heterogeneity in the effect sizes ($Q = 298.592$, $p < 0.0001$, $n = 45$). Forest plot showed that there was a clear outlier (article id 7 with production forest comparator) (Fig. 5) so we rerun the analysis without it. The overall effect size remained significantly different from zero ($d = 0.707$, $p < 0.0001$ 95% CI 0.366, 1.049). Sensitivity analysis was conducted
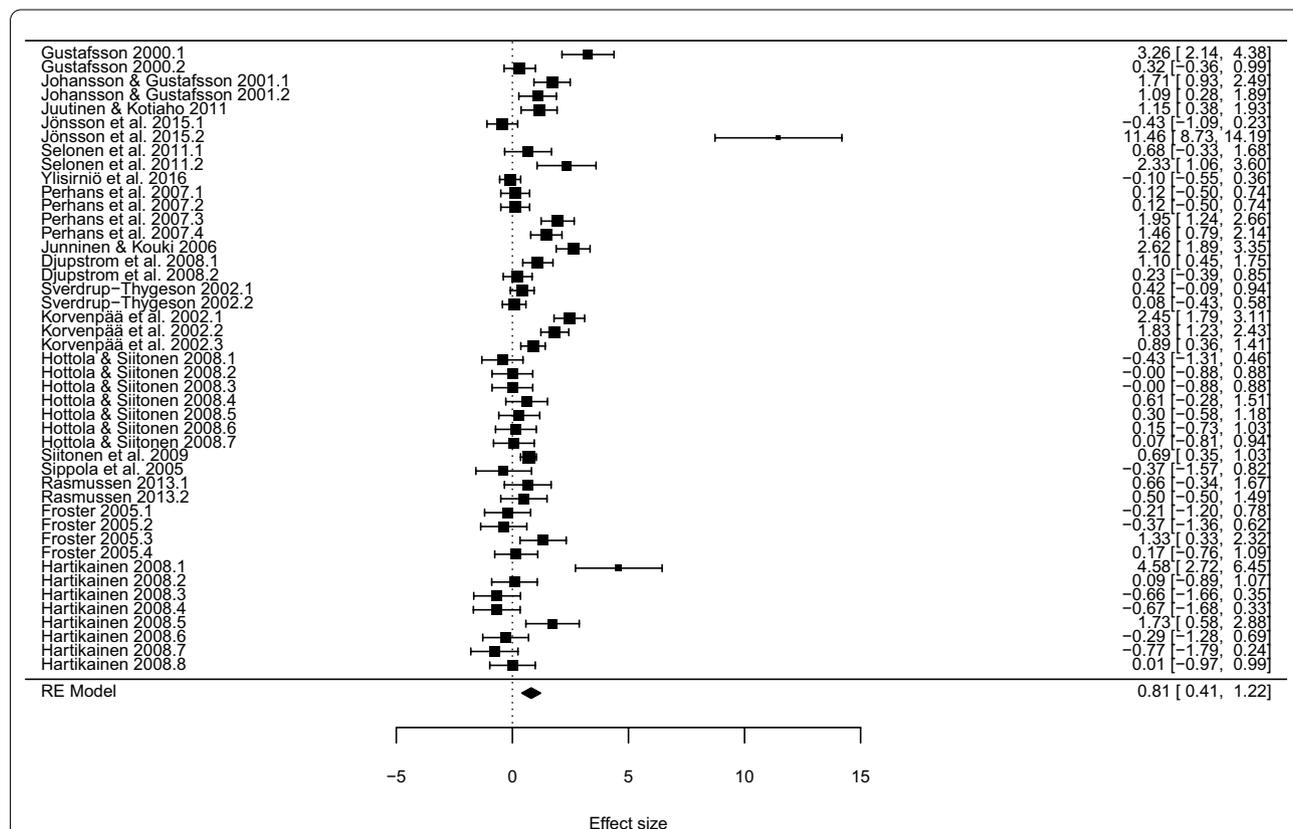
Fig. 5 Forest plot for the overall species richness between SPHPs and comparator forest types (production and natural forest). Effect sizes on the right side of zero mean that SPHP has more species than comparator forest. The grand mean noted by a diamond shows the summary effect of all the individual effect sizes shown on the right side of the figure. The error bars represent 95% confidence intervals

by excluding studies that had multiple treatments and thus, corrected sample sizes. Exclusion had no effect on the results: species richness remained significantly higher in SPHPs than in comparator forests. (d = 0.816, p = 0.0003, 95% CI 0.371, 1.26, n = 33). Similarly, heterogeneity remained significant (Q = 190.691, p < 0.0001, n = 33). Type of publication (grey or peer-reviewed literature), country or year when data was collected had no influence on effect sizes (QM = 1.762, p = 0.779, n = 45). Funnel plot did not show any obvious asymmetry (Additional file 9).

Further analysis showed that species richness in SPHPs does not differ significantly from natural forest but they are significantly more diverse than production forests (d = 1.0807, p = 0.04 0, 95% CI 0.252, 1.891, n = 45). Taxa had no effect on species richness (p > 0.05 for all). When interaction between forest type and taxa was studied, we found that indicator species (lichens and fungi) were significantly more common in SPHPs than in production forests but this was based only on one study (d = 10.813, p = 0.0001, 95% CI 7.085, 14.541). When we removed the outlier study (id 7) and reran the analysis, no impact of comparator, taxa or their interaction was

found (QM = 18.267, p = 0.076, n = 44). Subsets of data were used to test the influence of age of the comparator forest, forest owner (private or company; there was only one state owned forest, so it was excluded from the analysis), and intensity of management in the area surrounding the SPHP. There was no impact of the management intensity of the area surrounding the SPHPs on the effect sizes (QM = 0.494, p = 0.482, n = 26). Also, the owner of the forest (QM = 2.773, p = 0.096, n = 19) nor the age of the comparator forest (QM = 2.634, p = 0.104, n = 16) influenced species richness.

Individual abundance

The overall mean effect size for abundance showed that SPHPs have significantly more individuals than comparator areas (d = 1.911, p = 0.0003, 95% CI 0.868, 2.954, n = 30) (Fig. 6) but there was significant heterogeneity in the results (Q = 254.813, p < 0.0001). The effect remained significant after removal of studies that had multiple treatments and thus, corrected sample sizes (d = 2.101, p = 0.0007, 95% CI 0.891, 3.311, n = 22). The influence of year when data was collected, country and publication type (journal article or master's thesis) on effect sizes was
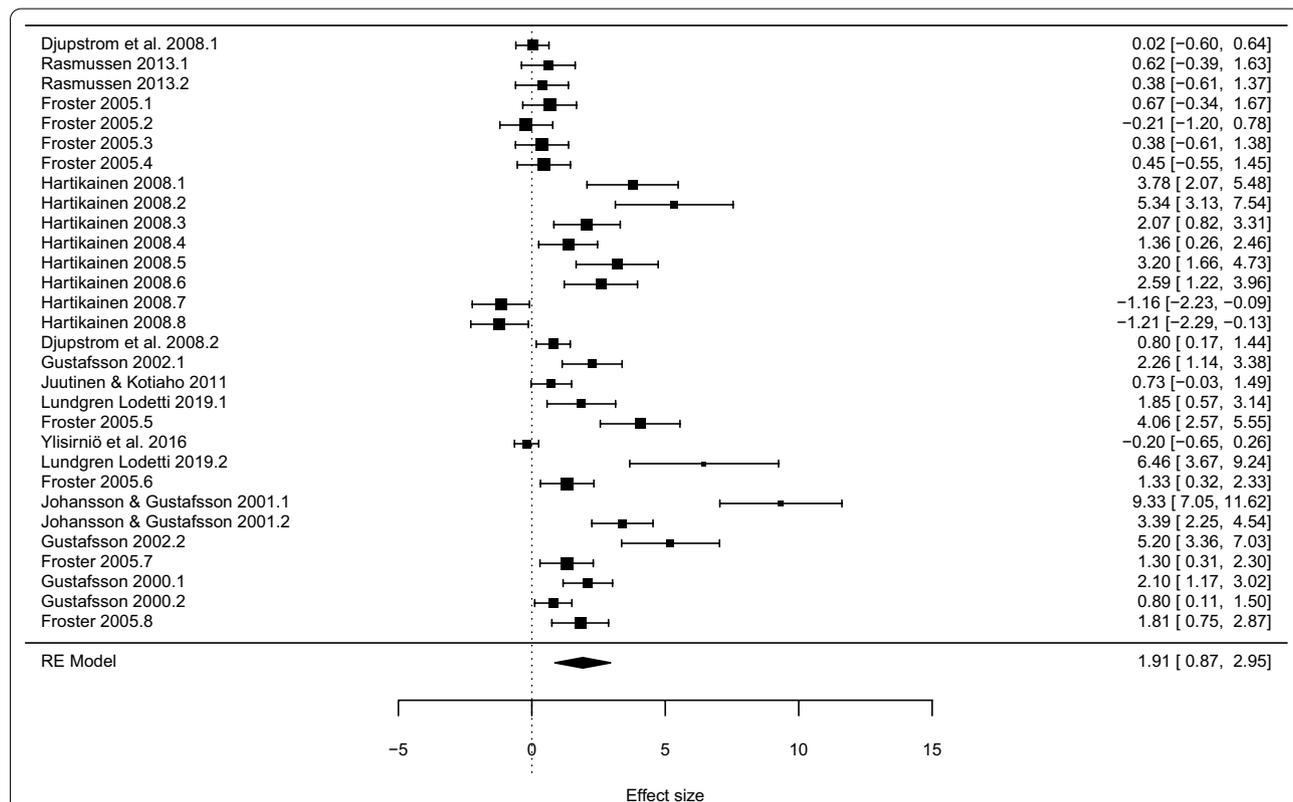
Fig. 6 Forest plot for individual abundance between SPHPs and comparator forest types (production and natural forest). Effect sizes on the right side of zero mean that SPHP has more individuals than comparator forest. The grand mean noted by a diamond shows the summary effect of all the individual effect sizes shown on the right side of the figure. The error bars represent 95% confidence intervals

tested but none of these were significant (QM = 1.486, p = 0.829, n = 30). Visual inspection of the funnel plot did not show an asymmetrical pattern. In addition, based on the results of the trim and fill-method no publication bias was detected (Additional file 9). When the influence of comparator type (natural or production forest) and taxa were tested, no differences in species abundance were found (QM = 7.216, p = 0.205, n = 30).

Deadwood

There was more deadwood overall in SPHPs than in controls (d = 0.625, p = 0.025, 95% CI: 0.079, 1.171, n = 26) (Fig. 7). There was significant heterogeneity in the effect sizes (Q = 882.700, p < 0.0001). When studies with imputed SDs were excluded, there was more deadwood in SPHPs than in controls, but the overall effect size was only marginally significant (d = 0.4795, p = 0.058, 95% CI − 0.017, 0.976). Effect sizes remained significantly heterogeneous (Q = 856.554, p < 0.0001). Similarly, the effect remained only marginally significant after exclusion of studies with corrected sample sizes (d = 0.963, p = 0.098, 95% CI − 0.1780.0271,

2.103, n = 16) but heterogeneity remained significant (Q = 137.765, p < 0.0001).

Effect sizes were not influenced by country, publication type and data collection year when all the studies were included (QM = 4.054, p = 0.399, n = 26). When studies with imputed SDs were removed, study year and publication type influenced the results (QM = 9.532, p = 0.049, n = 24). This was caused mainly by one study that was published in a peer-reviewed journal and had a relatively large positive effect size (d = 4.822).

No publication bias was visually detected, and this was confirmed by the results of the trim and fill-test (Additional file 9).

Comparator had significant influence on effect sizes (QM = 9.901, p = 0.002, n = 26)) but heterogeneity remained significant (Q = 232.258, p < 0.0001) indicating that there are other moderators influencing effect sizes. Based on the analysis, deadwood volumes were significantly higher in SPHPs compared to production forests (d = 1.564, p = 0.002, 95% CI 0.590, 2.539, n = 26). The influence of age of production forest was tested but no effect was found (QM = 0.954, p = 0.323, n = 9). When studies with imputed SDs were excluded, the results
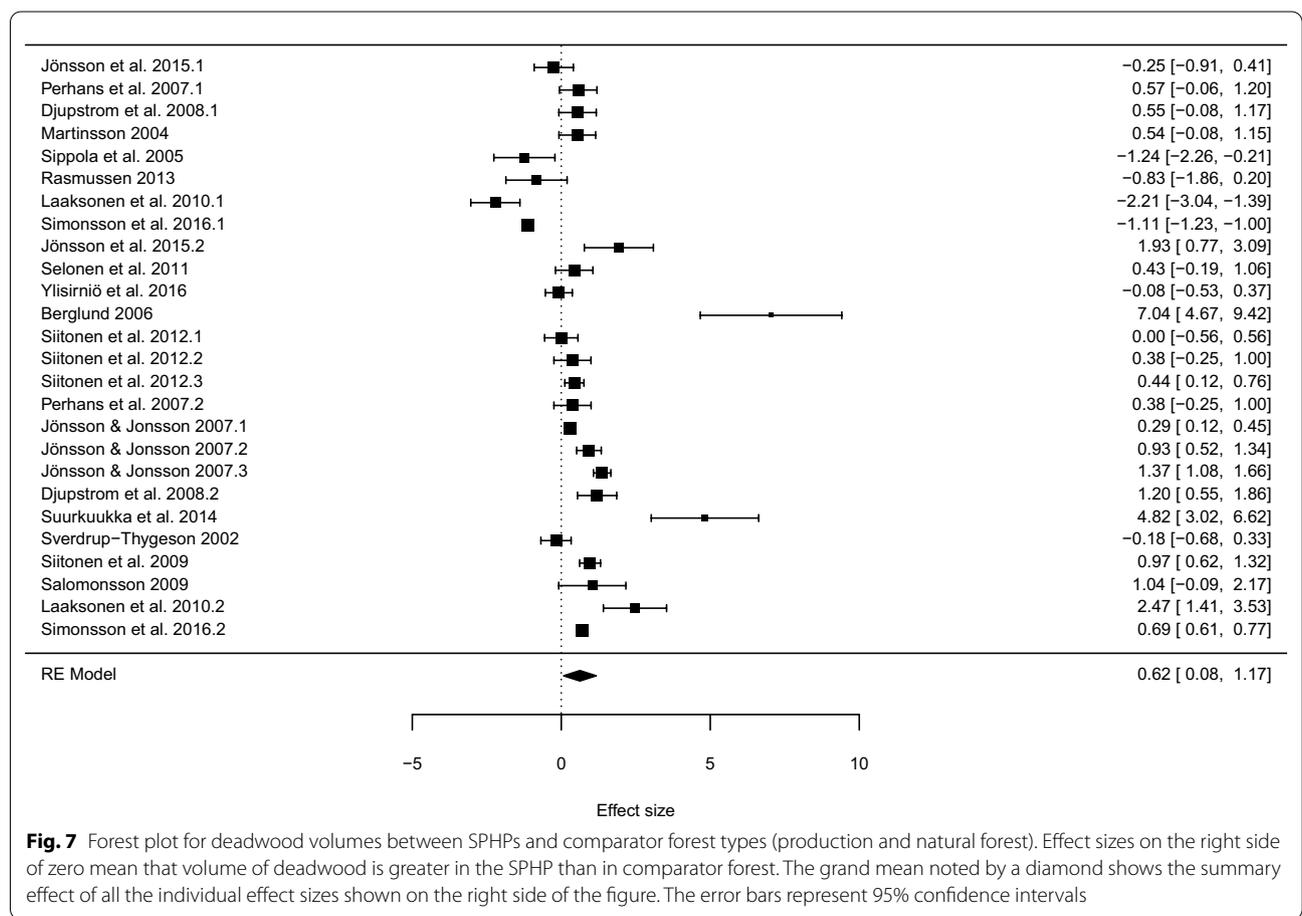
Fig. 7 Forest plot for deadwood volumes between SPHPs and comparator forest types (production and natural forest). Effect sizes on the right side of zero mean that volume of deadwood is greater in the SPHP than in comparator forest. The grand mean noted by a diamond shows the summary effect of all the individual effect sizes shown on the right side of the figure. The error bars represent 95% confidence intervals

remained similar. Deadwood volumes were significantly higher in SPHPs compared to production forests (d = 1.491, p = 0.0005, 95% CI 0.647, 2.336, n = 24). The excluded studies did not record the age of production forest.

Studies with high risk of bias
When studies with high risk of bias (n = 3) were excluded from the dataset, the overall effect size was only marginally significant (d = 0.612, p = 0.06, n = 23). Also, age of the comparator forest had marginally significant impact on the effect sizes (d = − 0.009, p = 0.061, 95% CI − 0.019, 0.0004, n = 7). Otherwise, the results were similar to the analyses where studies with high risk were not removed. Results without high-risk studies are included as Additional file 10.

Assemblage
Similarity of species composition could be statistically assessed in 26 studies. Further 18 studies provided some information about species composition (Additional file 11).

A mean of shared species between SPHPs and controls was calculated to compare similarity of species composition. There were 14 studies on production forests and 12 on natural forests that provided suitable data for comparison. On average, production forests and SPHPs shared approximately 58% of species and natural forests and SPHPs 48% of species after the values were standardized (Fig. 8). This represents significant differences in community composition regarding both forest types.

Evidence of effects
Overall, the evidence supports using SPHPs, specifically WKHs, to maintain and increase biodiversity in production forests. There were significantly more species in

SPHPs than in production forests concurring with earlier systematic review on WKHs [35]. Also, dead wood volumes were higher in SPHPs than in production forests. The larger volume of dead wood in SPHPs is not surprising since removal of dead wood in SPHPs is usually forbidden, while in production forests dying and dead wood have mainly been harvested in the past. However, nowadays it is often recommended to leave some deadwood also in production forests (for example [67,68]). Increasing the amounts of deadwood have been shown to have a positive effect on saproxylic insects and fungi (i.e. species dependent on dead wood) [69], which indicates positive reinforcement of biodiversity benefits of SPHPs.

Contrary to expectations, management intensity of the area surrounding SPHPs did not have an influence on species richness. There have been studies indicating that even in the case of protected areas, the management of the surroundings can affect species diversity (for example [70] on birds). However, these effects can vary between species groups, and most of the studies in this review were about plants and arthropods. In addition, it must be noted that in many studies the management intensity of the area surrounding SPHPs was not reported, and none of the studies reported light management intensity. Hence, the lack of effect can also mean that taxa react similarly to both moderate and high intensity management. It is also possible that SPHPs are suffering from extinction debt, i.e. even though the species are still present in SPHPs, they can become extinct in the future because isolation of the patch can lead to lack of gene flow and viable populations in long term [33, 34]. However, long-term studies are needed to explore the phenomenon. Owner of SPHP did not have influence on species richness.

No differences in species richness between SPHPs and natural forest were noticed concurring with the results of Timonen et al. [9]. When species assemblages were studied, differences between SPHPs and production forests as well as SPHPs and natural forests were found. This means that although SPHPs harbour similar number of species than natural forests, they are not 'mini natural forests' in terms of species composition. This may result partly from the definition criteria of WKHs (all SPHPs analysed here were WKHs). In some cases, the criteria include indicator species, which means these species inevitably exist in the WKH. In addition, WKHs are already by definition so called biodiversity hotspots where large number of species occur within small area. In production forests and natural forests, species are distributed within a larger area. The same species may not occur in the randomly chosen study plot in the comparator forest, but this does not mean that the species could not occur in some other

Fig. 8 The mean proportion of shared species between SPHPs and control areas with 95% confidence intervals. Data were standardized by the total number of species recorded in SPHPs (the number of species = 1)
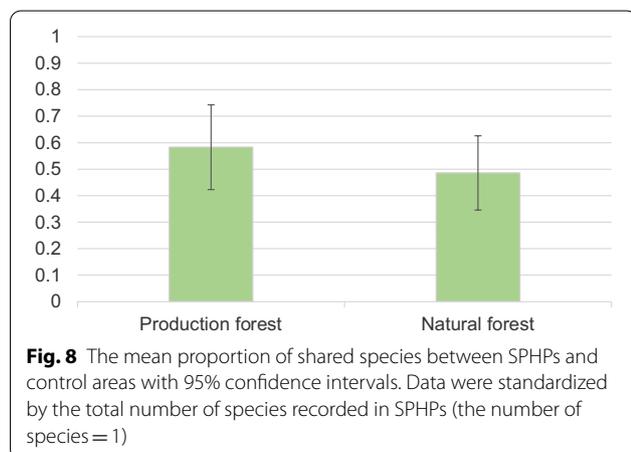
location within the same forest. Also, the differences in species assemblages probably reflect differences in habitat requirements, e.g. in terms of light, humidity and exposure to edge effects. However, more studies are needed, especially regarding species assemblages in SPHPs and neighbouring production forests, to have conclusive evidence on the causal links that lead to certain composition of species.

Species abundance in SPHPs was higher overall than in natural forests and production forests but not when the effect was separated per comparator forest type. The explanation can be related to special attributes of SPHPs. As the evidence shows, they can harbour more resources, like dead wood, and therefore they can host more individuals dependent on those resources.

Magnitude of effects

Where SPHPs had significant or even marginally significant impact on species richness or abundance, the effect sizes were in most cases large. Usually, effect size of 0.2 is considered a small effect, $d = 0.5$ an intermediate effect and $d = 0.8$ a large effect [71]. In our results, effect sizes were commonly above one, and large effect sizes indicate a strong response from the studied groups. Considering that most of the observational studies had aimed to minimise bias from environmental variation across study sites, we are confident that the large effects found in this systematic review are representative of true effects in nature for the studied species groups.

Review limitations

Even though different sources were searched widely, it is likely that some relevant articles were not found. Because the languages and geographical scope of the review did not fully overlap, there may be relevant articles written in Norwegian, Estonian, Latvian, Lithuanian, or French (Canada) that were not included in this review. Searching every organisational website or database that may include relevant material is impossible, but the search in this review was comprehensive and no publication bias was detected. It is worth noticing that almost 40% of the articles included in this review were found outside the citation databases (Scopus and WoS) and almost 30% were grey literature.

Studies included in this review were not 100% comparable with each other but restricting the eligibility criteria even more could have resulted in a situation where quantitative analysis was not possible. For example, the comparators within one comparator class (natural or production forest) differed. The age of production forests varied, and natural forests were both protected and unprotected. However, these differences reflect variability in nature, and hence natural sources of heterogeneity.

Comparability of production forests was enhanced by excluding articles where the production forest was already, before the study, known to have low or high conservation values. These were, for example, areas not accepted in the METSO program and areas receiving environmental subsidies.

Geographical distribution of the articles included in this review was limited. There were many articles from Finland and Sweden, few from Norway and only one from Latvia. Since the concept of WKH is widely used in all the Baltic Countries and to some extent also in Western Russia, the lack of studies from these countries is unexpected. In Latvia there has been a lot of research on WKHs, but the focus has been on describing the species in WKHs or comparing different types of WKHs (for example [63,64]). Also, many forests in Latvia are temperate and studies conducted in non-boreal forests were excluded from this review.

The results can also be affected by the studied taxa. Most of the studies included in this review were concentrating on small, sedentary, and poorly distributing organisms like insects, plants, and fungi. The importance of SPHPs for these small and local organisms is clear but based on this review the importance of SPHPs for example for mammals and birds is not known.

Not all studies included in narrative synthesis had suitable data for quantitative analysis and hence the quantitative synthesis could not be based on the whole knowledge base. Reporting on the effect modifiers was also inadequate in many articles. This naturally affected meta-analyses when the influence of effect modifiers could not be analysed. Especially age of the comparator forest and the type of management of the area surrounding SPHPs would have been interesting effect moderators to analyse with more data. The lack of studies comparing WKHs with differently managed surroundings was highlighted already in a former systematic review by Timonen et al. [35].

Review conclusions
Implication for policy/management

There were more species, individuals, and dead wood in WKHs than in production forests which indicates that setting aside SPHPs helps to protect biodiversity in production forests. However, species assemblages differed between SPHPs and natural forests meaning that SPHPs cannot be considered as a substitute for large natural forests like national parks and reserves. But SPHPs do complement the protected area network and may increase connectivity between larger reserves by providing habitats for species with restricted dispersal abilities.

Most of the studies concentrated on rare/indicator/signal species and data extraction were also targeted

towards these species. Therefore, the results cannot be straightforwardly generalized to all species, not even to all species of the most studied species groups. However, because the purpose of setting aside small habitat patches is to protect especially rare species, the results indicate the success of achieving this target.

Also, clear majority of the studies were about WKHs and therefore, the results cannot be generalized directly to other types of small protected habitat patches when they do not share similar attributes with WKHs. The geographical distribution of the included studies should also be considered when generalizing the results. Most of the studies were conducted in Finland and Sweden, and even between these two countries the definitions, sizes, and protection status of WKHs differs.

Implication for research

Based on this systematic review, there are knowledge gaps regarding geographical distribution, interventions and studied organisms. More research on SPHPs in general is needed, and especially research comparing SPHPs and production forests or natural forests is needed in Baltic Countries and Russia, and to some extent also in Norway. Alongside WKHs, future research should concentrate on other small protected habitat patches, such as voluntarily protected areas.

The results of this review indicate that importance of SPHPs to mammals and birds is not known and more studies concentrating on these species is needed. In addition, replicating past studies in the future would offer more information about the ability of SPHPs to maintain biodiversity in the long run. To enhance reliability of the individual studies and meta-analyses in the future, it is important to report more in detail about potential effect modifiers and reasons for heterogeneity. Especially reporting the management intensity of the area surrounding the SPHPs is important for achieving information on the ability of SPHPs to maintain biodiversity within different landscape matrices.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13750-020-00216-6.

Additional file 1. ROSES form for systematic review reports.

Additional file 2. Corrected test list. The test list modified after publication of the review protocol.

Additional file 3. Corrected performance of the search string. The performance of the search string after modifications of the search string and test list.

Additional file 4. Search strings. Full search strings in English and Russian, and simplified search strings for Google and Google Scholar in English, Russian, Swedish and Finnish.

Additional file 5. Search strategy and results. A list of all sources searched with information on search dates, languages, search strings/terms, search settings/restrictions, the number of hits received, and the number of articles included on each stage of the screening.

Additional file 6. List of articles included by title/abstract. There are different sheets for articles included by full text, articles included by full text with independent data (articles in narrative synthesis), articles excluded by full text (with reasons for exclusion) and unretrievable articles.

Additional file 7. Metadata and validity assessment. File includes metadata of different studies, data on effect size modifiers and validity assessment for each study.

Additional file 8. R script for meta-analyses.

Additional file 9. Funnel plots.

Additional file 10. Results of the analysis of dead wood volume without high risk studies.

Additional file 11. Summary information about similarity of species composition.

Authors' contributions
This review was written by MH, SS and AJ. AU and TS built and tested the final search string. AU, TS, MH, and AJ conducted the searches. MH, SS and AJ screened the articles. Data extraction and critical appraisal was conducted by AJ and MH. AJ conducted the narrative synthesis and SS the quantitative synthesis. MH, AJ, and SS wrote the first version of the manuscript. MM and PP contributed to the subsequent versions of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its additional files.

Ethics approval and consent to participate
Not applicable. Consent for publication. Not applicable.

Consent for publication
Not applicable.

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

Author details
[1] Department of Biological and Environmental Science, University of Jyväskylä, P.O.Box 35, 40014 Jyväskylä, Finland. [2] Metsäteho Oy, Vernissakatu 1, 01300 Vantaa, Finland. [3] Helsinki University Library, University of Helsinki, Fabianinkatu 30, 00014 Helsinki, Finland. [4] Helsinki University Library, Viikki Campus, Viikinkaari 11 A, 00014 Helsinki, Finland. [5] School of Resource Wisdom, University of Jyväskylä, P.O.Box 35, 40014 Jyväskylä, Finland. [6] Department of Forest Sciences, University of Helsinki, Latokartanonkaari 7, 00014 Helsinki, Finland.

Häkkilä *et al. Environ Evid*   (2021) 10:2

Page 19 of 20

References
1. FAO. Global Forest Resources Assessment 2005. Progress towards sustainable forest management. 2006.
2. Lindenmayer D, Franklin JF. Conserving forest biodiversity: a comprehensive multiscaled approach. Washington: Island Press; 2002. p. 351.
3. Naughton-Treves L, Holland MB, Brandon K. The role of protected areas in conserving biodiversity and sustaining local livelihoods. Annu Rev Environ Resour. 2005;30(1):219–52.
4. Gaston KJ, Jackson SF, Cantú-Salazar L, Cruz G, Jackson SE, Cantfi-Salazar L, et al. The ecological performance of protected areas. Annu Rev Ecol Evol Syst. 2008;39(39):93–113.
5. Watson JEM, Dudley N, Segan DB, Hockings M. The performance and potential of protected areas. Nature. 2014;515(7525):67–73.
6. Scott JM, Davis FW, Mcghie RG, Wright RG, Groves C, Estes J. Nature reserves: Do they capture the full range of America´s biological diversity? Ecol Appl. 2001;11(August):999–1007.
7. Gustafsson L, Baker SC, Bauhus J, Beese WJ, Brodie A, Kouki J, et al. Retention forestry to maintain multifunctional forests: a world perspective. Bioscience. 2012;62(7):633–45.
8. Vanha-Majamaa I, Jalonen J. Green tree retention in fennoscandian forestry. Scand J For Res. 2001;16:79–90.
9. Timonen J, Siitonen J, Gustafsson L, Kotiaho JS, Stokland JN, Sverdrup-Thygeson A, et al. Woodland key habitats in northern Europe: concepts, inventory and protection. Scand J For Res. 2010;25(4):309–24.
10. Nitare J, Norén M. Nyckelbiotoper kartlåggs i nytt projekt vid Skogsstyrelsen [Key biotopes are mapped in a new project at Skogsstyrelsen]. Sven Bot Tidskr. 1992;86:219–26.
11. Ericsson TS, Berglund H, Östlund L. History and forest biodiversity of woodland key habitats in south boreal Sweden. Biol Conserv. 2005;122(2):289–303.
12. Akatova T, Bibin A, Grabenko E, Zagurnaâ Û. Ключевые биотопы Эксплуатируемых лесов - Краснодарского края и Республики Адыгея (Северо-кавказский горный регион) ("Key biotopes in exploited forests - Krasnodarsk Krai and Republic of Adygea (North Caucasian mountain region)")["Key biotopes in exploited. Устойчивое Лесопользование. 2016;3(47):29–35.
13. Raj EA, Torhov SV, N.V. B, Rykova SÛ, Amosov PN, Korepanov VI, et al. Ключевые биотопы лесных экосистем Архангельской области и рекомендации по их охране ["Key biotopes of forest ecosystems in Arkhangelsk Oblast and recommendations for their conservation]. WWF, Russia, Arkhangelsk; 2008. 30 p.
14. Suomen Säädöskokoelma. Laki metsälain muuttamisesta 1085/2013. [Forest Act 1093/1996, amendments up to 567/2014 included. 2013. http://www.finlex.fi/fi/laki/kaannokset/1996/en19961093.pdf
15. Aasaaren Ø, Sverdrup-Thygeson A. Nøkkelbiotoper i skogen [Key biotopes in forests]. Oslo: NORSKOG; 1994.
16. Meriluoto M, Soininen T. Metsäluonnon Arvokkaat Elinympäristöt [Valuable habitats of forest nature]. Metsälehti Kustannus: Tapio; 1998.
17. Norén M, Nitare J, Larsson A, Hultgren B, Bergengren I. Handbok för inventering av nyckelbiotoper [Handbook for key habitat inventory]. Jönköping: Skogsstyrelsen; 2002.
18. Prieditis N. Evaluation frameworks and conservation system of Latvian forests. Biodivers Conserv. 2002;11(8):1361–75.
19. Andersson L, Kriukelis R, Skuja S. Woodland key habitat inventory in Lithuania Vilnius: Lithuanian Forest Inventory and Management Institute Kaunas; Linköping: Regional Forestry Board of Östra Götaland; 2005.
20. Gjerde I, Sætersdal M, Blom HH. Complementary Hotspot Inventory–a method for identification of important areas for biodiversity at the forest stand level. Biol Conserv. 2007;137(4):549–57.
21. Mäntymaa E, Juutinen A, Mönkkönen M, Svento R. Participation and compensation claims in voluntary forest conservation: A case of privately owned forests in Finland. For Policy Econ. 2009;11(7):498–507.
22. Widman U. Exploring the role of public-private partnerships in forest protection. Sustainability. 2016;8(5):496.
23. Storrank B. Nordiska Arbetspapper - Frivilligt skydd av skog i Finland, Sverige och Norge ("Nordic Working Paper - Voluntary Protection of Forests in Finland, Sweden and Norway"). 2018. http://norden.diva-portal.org/smash/get/diva2:1190782/FULLTEXT01.pdf
24. USDA. National Report on Sustainable Forests. 2011. https://www.fs.fed.us/research/sustain/docs/national-reports/2010/2010-sustainability-report.pdf
25. Põllumäe P, Korjus H, Paluots T. Management motives of Estonian private forest owners. For Policy Econ. 2014;42:8–14.
26. Pivoriūnas A, Lazdinis M. Needs of private forest owners in the context of changing political systems: Lithuania as a case study. Small-scale For Econ Manag Policy. 2004;3(2):191–202.
27. FSC. Forest Stewardship Council. International Generic Indicators FSC-STD-01–004 V1–0 EN. Bonn, Germany; 2015.
28. PEFC. Programme for the Endorsement of Forest Certification. PEFC ST 1003:2018. Sustainable Forest Management–requirements. Geneva; 2018.
29. FSC. Forest Stewardship Council. National Boreal Standard. Canada. 2004.
30. FSC. FSC Standard for Finland V1–1. 2011.
31. SFI. Sustainable Forestry Initiative SFI 2015–2019 Forest management standard. Washington D.C.: Sustainable Forestry Initiative Inc; 2015.
32. Hanski I. The shrinking world: Ecological consequences of habitat loss. In: Kinne O, editor. Excellence in Ecology. Oldendorf: International Ecological Institute; 2005.
33. Berglund H, Jonsson BG. Verifying an extinction debt among lichens and fungi in northern swedish boreal forests. Conserv Biol. 2005;19(2):338–48.
34. Ylisirniö A-L, Mönkkönen M, Hallikainen V, Ranta-Maunus V, Kouki J. Woodland key habitats in preserving polypore diversity in boreal forests: effects of patch size, stand structure and microclimate. For Ecol Manage. 2016;373:138–48.
35. Timonen J, Gustafsson L, Kotiaho JS, Mönkkönen M. Hotspots in cold climate: conservation value of woodland key habitats in boreal forests. Biol Conserv. 2011;144(8):2061–7.
36. Häkkilä M, Savilaakso S, Johansson A, Sandgren T, Uusitalo A, Mönkkönen M, Puttonen P. Do small protected habitat patches within boreal production forests provide value for biodiversity conservation? A systematic review protocol. Environm Evidence. 2019;8(1):30.
37. Keenan RJ, Reams GA, Achard F, de Freitas JV, Grainger A, Lindquist E. Dynamics of global forest area: Results from the FAO global forest resources assessment 2015. For Ecol Manage. 2015;352:9–20.
38. Livoreil B, Glanville J, Haddaway NR, Bayliss H, Bethel A, de Lachapelle FF, et al. Systematic searching for environmental evidence using multiple tools and sources. Environ Evid. 2017;6(1):23.
39. Cheng SH, Augustin C, Bethel A, Gill D, Anzaroot S, Brun J, DeWilde B, Minnich RC, Garside R, Masuda YJ, Miller DC, Wilkie D, Wongbusarakum S, McKinnon MC. Using machine learning to advance synthesis and use of conservation and environmental evidence 2015. Conserv Biol. 2018;32(4):762–4.
40. Frampton GK, Livoreil B, Petrokofsky G. Eligibility screening in evidence synthesis of environmental management topics. Environ Evid. 2017;6(1):27.
41. Rohatgi A. WebPlotDigitizer, version 4.2. https://automeris.io/WebPlotDigitizer. 2019.
42. Rosenberg MS, Rothstein HR, Gurevitch J. Effect sizes, conventional choices and calculations. In Koricheva J, Gurevitch J, Mengersen K Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press 2013
43. Gleser JJ, Olkin I, Stochastically dependent effect sizes. In: Cooper H, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Russell Sage Foundation; 2009.
44. Lajeunisse MJ, Recovering missing or partial data from studies. In Koricheva J, Gurevitch J & Mengersen K, . Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press; 2013.
45. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2000, Vienna. URL https://www.R-project.org/.
46. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software 2010; 36(3): 1–48. https://www.jstatsoft.org/v036/i03.
47. Cochran WG. Some methods for strengthening the common x2 tests. Biometrics. 1954;10:417–51.
48. Nichols E, Larsen T, Spector S, Davis AL, Escobar F, Favila M, Vulinec K. The Scarabaeinae research network. Global dung beetle response to tropical forest modification and fragmentation: a quantitative literature review and meta-analysis. Biol Conservat. 2007;137(1):1–19.
49. Haddaway NR, Macura B, Whaley P, Pullin AS. ROSES flow diagram for systematic reviews. Version. 2017. https://doi.org/10.6084/m9.figshare.5897389.

50. Siitonen J, Penttilä R, Ihalainen A. METSO-ohjelman uusien pysyvien ja määräaikaisten suojelualueiden ekologinen laatu Uudenmaan alueella. Metsätieteen aikakauskirja. 2012;4:259–83.

51. Simonsson P, Östlund L, Gustafsson L. Conservation values of certified-driven voluntary forest set-asides. For Ecol Manage. 2016;375:249–58.

52. Korvenpää T, Lehesvirta T, Salpakivi-Salomaa P. Pienvesien avainbiotoopit tärkeitä harvinaisille sammalille. Luonnon tutkija. 2002;106:5.

53. Hottola J, Siitonen J. Significance of woodland key habitats for polypore diversity and red-listed species in boreal forests. Biodiv Conservat. 2008;17(11):2559–77.

54. Siitonen J, Hottola J, Immonen A. Differences in stand characteristics between brook-side key habitats and managed forests in southern Finland. Silva Fennica. 2009. https://doi.org/10.14214/sf.216.

55. Selonen V, Mussaari M, Toivanen T, Kotiaho JS. The conservation potential of brook-side key habitats in managed boreal forests. Silva Fennica 2011; 45.

56. Suurkuukka H, Virtanen R, Suorsa V, Soininen J, Paasivirta L, Muotka T. Woodland key habitats and stream biodiversity: Does small-scale terrestrial conservation enhance the protection of stream biota? Biol Cons. 2014;170:10–9.

57. Hartikainen H. Metsälain (1093/1996) 10 §: n määrittelemien lehtojen merkitys putkilokasvilajiston monimuotoisuuden säilymiselle. Master's thesis, University of Jyväskylä, Faculty of Science, Department of Ecological and Environmental Science, 2008.

58. Lundgren Lodetti M. Log-diversity and abundance of dead wood dependent bracket fungi and bryophytes, a comparison between spruce woodland key habitats and adjacent production forests in Östergötland, Sweden. Master's thesis, Linköping University,. Department of Physics, Chemistry and Biology, 2019.

59. Froster A. Occurrence of signal species in woodland key habitats, nature reserves and production forest. Master's thesis. Department of Conservation Biology. Swedish University of Agriculture, Uppsala, 2005, 32 p

60. Perhans K, Gustafsson L, Jonsson F, Nordin U, Weibull H. Bryophytes and lichens in different types of forest set-asides in boreal Sweden. For Ecol Manage. 2007;242(2–3):374–90.

61. Djupström LB, Weslien J, Schroeder LM. Dead wood and saproxylic beetles in set-aside and non set-aside forests in a boreal region. For Ecol Manage. 2008;255(8–9):3340–50.

62. Laaksonen M, Murdoch K, Siitonen J, Várkonyi G. Habitat associations of Agathidium pulchellum, an endangered old-growth forest beetle species living on slime moulds. J Insect Conserv. 2010;14(1):89–98.

63. Jönsson M, Ruete A, Gunnarsson U, Kellner O & Snäll T. Övervakning av värdefulla skogsbiotoper: en utvärdering av extensivmetoden efter 10 år. SLU report, 2015.

64. Gjerde I, Sætersdal M, Blom HH. Complementary hotspot inventory–a method for identification of important areas for biodiversity at the forest stand level. Biol Cons. 2007;137(4):549–57.

65. Putna S, Mežaka A. Preferences of epiphytic bryophytes for forest stand and substrate in North-East Latvia. Folia Cryptogamica Estonica. 2014;51:75–83.

66. Jönsson MT, Jonsson BG. Assessing coarse woody debris in Swedish woodland key habitats: Implications for conservation and management. For Ecol Manage. 2007;242(2–3):363–73.

67. PEFC Finland. Criteria for PEFC Certification. PEFC FI 1002:2014.

68. FSC Sweden. Swedish FSC Standard for Forest Certification Including SLIMF indicators. FSC-STD-SWE-02–04–2010, Sweden Natural, Plantations and SLIMF EN.

69. Sandström J, Bernes C, Junninen K, Lõhmus A, Macdonald E, Müller J & Jonsson BG. Impacts of dead wood manipulation on the biodiversity of temperate and boreal forests. A systematic review. Journal of Applied Ecology 2019; 56:1770–1781.

70. Häkkilä M, Le Tortorec E, Brotons L, Rajasärkkä A, Tornberg R, Mönkkönen M. Degradation in landscape matrix has diverse impacts on diversity in protected areas. PLoS ONE. 2017;12(9):e0184792.

71. Cohen J. Statistical power analysis for the behavioral sciences. New York: Routledge Academic; 1988.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.