Collaboration for Environmental Evidence
Library

CEE review 09-015

WHAT IS THE IMPACT OF 'LIMING' OF STREAMS AND RIVERS ON THE ABUNDANCE AND DIVERSITY OF FISH AND INVERTEBRATE POPULATIONS?

Systematic Review

MANT, R. [1]**, JONES, D.** [1]**, REYNOLDS, B.** [2]**, ORMEROD, S.** [3] **& PULLIN, A.S.**[1]

[1] School of Environment, Natural Resources & Geography, Bangor University, Deiniol Road, Bangor, LL57 2UW, UK

[2] Centre for Ecology & Hydrology, Environment Centre Wales, Deiniol Road, Bangor, LL57 2UW , UK

[3] Cardiff School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UK.

Correspondence: rebecca.mant@cantab.net
Telephone: +44 (0)1248 382953
Fax:

Draft protocol published on website: 10 July 2009 - Final protocol published on website:21 July 2010 - Draft review published on website: 7 July 2011- Final review published on website: 6 December 2011.

Cite as: Mant, R., Jones, D., Reynolds, B., Ormerod, S. & Pullin, A. 2011. What is the impact of liming of streams and rivers on the abundance and diversity of fish and invertebrates? CEE review 09-015 (SR76). Collaboration for Environmental Evidence: www.environmentalevidence.org/SR76.html.

Summary

1. Background
Calcium carbonate has been applied extensively to mitigate the impacts of surface-water acidification caused by 'acid rain' but there are still uncertainties about the effects on fish and invertebrates. For the first time, this systematic review summarises the best available evidence on the impact of liming on invertebrate assemblages and fish populations in rivers.

2. Methods
A systematic search for relevant articles used terms describing rivers, the biota of interest and the intervention. All retrieved articles were scanned for relevance using specified inclusion criteria. Included articles were appraised critically; study methods were identified along with the presence of confounding factors. The main findings of the studies were extracted individually for each outcome of interest (fish abundance, fish diversity, invertebrate abundance, etc.) and the response calculated as the ratio of diversity or abundance in the limed sample to that in the unlimed control. A random effects meta-analysis was carried out on the log of the ratios.

3. Main results
Thirty-four independent studies met the stringent selection criteria, from which data were subsequently extracted. Only a minority had both control and baseline data (a Before, After, Control, Impact, BACI study design). Hence, there was a risk of bias due to differing baselines between the treatment and control groups and factors other than liming changing during the course of the study. All studies were included in the analysis but sensitivity of the results to the study design was tested.

Over all studies liming increased fish abundance, with mean response ratio of 1.7 (C.I 1.3 - 2.1). However, the effect varied between studies and fish abundance was predicted to decrease in around one fifth of limed rivers; the interval predicted to contain 95% of true study effects was 0.7 to 4.3. The variability was partially explained by the average duration of liming treatment and where salmon and trout occurred together they were impacted differently. There were no significant differences between studies of different experimental design.

Liming had no effect on invertebrate abundance over all studies. Including only BACI designed studies, liming on average reduced invertebrate abundance (mean response ratio = 0.78, CI= 0.61 – 0.99). The mean effect of liming on the number of invertebrate taxa was a significant increase (mean effect (response ratio) = 1.17, CI=1.03-1.33). However, again the effect varied significantly between studies and the prediction interval overlapped no effect. Effect of liming on acid sensitive invertebrate abundance was positive (mean response ratio = 1.96, CI=1.12-3.44), however, the prediction interval still overlapped no effect. There was a significant increase in the number of acid sensitive invertebrate taxa after liming (mean effect= 2.58, CI=1.65-4.02) largely reflecting just one study.

5. Conclusions
Liming is usually an effective intervention for increasing fish populations in streams and rivers affected by anthropogenic acidification but cannot be guaranteed in all situations. Positive effects are more likely in longer-term applications and on salmon, although again this is not guaranteed. Liming generally increased the abundance and taxonomic richness of

acid sensitive invertebrates but effects were variable and for all invertebrate taxa combined liming may decrease abundance.

Main Text

1. Background

"Acid rain" and the associated acidification of waterways first became a widespread environmental concern in the 1970's (Menz and Seip 2004). Sulphur dioxide and nitrogen oxides released during the burning of combustible waste and fossil fuels react with water in the atmosphere to produce sulphuric and nitric acid, reducing the pH of rainfall and creating 'acid rain' (Singh and Agrawal 2008). Acid rain has caused multiple problems including damage to limestone buildings (Chestnut and Mills 2005) and forest die-back (Likens et al. 1996). In the aquatic environment, acidification of waterways frequently causes changes in ecosystems including a reduction in fish and invertebrate numbers and diversity (Schindler et al. 1985, Moiseenko 2005). Acid rain can cause increased mobilisation of aluminium and hence, increased aluminium in waterways (Singh and Agrawal 2008) which can in turn harm fish (Liebich et al. 2011). Acidification has had a particularly large impact on salmon populations. In Norway, acid rain is thought to have caused the eradication of salmon from the two most southern counties by 1995 (Sandøy and Langåker 2001) and in Nova Scotia, Canada, to have caused a 50% reduction in salmon production in the 1980s (Watt 1987).

In order to mitigate the problems of 'acid rain', including acidification of waterways, considerable efforts have been made to reduce industrial emissions since the 1980's. The emissions of sulphur have been successfully reduced through various legislations, both in Europe and North America, reducing the (non-marine) sulphate within water bodies (Reynolds et al. 2004, Evans et al. 2001). However, nitrous oxide emissions have not decreased to the same extent as sulphate emissions (Fowler et al. 2007); levels of nitrate within surface waters have remained relatively unchanged in Wales (Reynolds et al. 2004) and responded variably throughout Europe (Evans et al. 2001). Overall, there are still areas where the acid load outstrips the soils neutralizing capacity (Matejko et al. 2009). Even where decreased deposition has caused the mean pH to rise, periods of high river flow can be associated with acid episodes (Evans et al. 2008). Additionally, as industrialisation has spread from Europe into South Asia, acid rain has become an increasing problem in other parts of the world (Kuylenstierna et al. 2001).

Despite the reductions in emissions, biological recovery in watercourses has been variable (Ormerod and Durance 2009) and acid episodes may be restricting biological recovery (Kowalik et al. 2007). As removing the cause of acidification by reducing emissions has not always allowed full recovery and taken time to implement, symptomatic mitigation strategies have also been used. Land use can impact on the effects of acidification and changes in land use (including conifer forestry) may alter the impact of acidification (Nisbet 2001). However, one of the most widespread mitigation techniques is the addition of calcium carbonate (usually in the form of crushed or powdered limestone) to a watercourse (Clair and Hindar 2005). The application of limestone, commonly termed 'liming', is undertaken to raise the pH of both rivers and lakes within a watercourse, often at the same time. However, as the ecology of lakes and rivers are different, the impact of liming may differ. In order to raise the pH of rivers, limestone has been applied using a wide variety of different methods and to different parts of a watercourse including directly into rivers, into lakes upstream of rivers and to the catchment areas of rivers (Donnelly et al. 2003). The details of the different liming methods have been set out in detail in Donnelly et al. (2003) and Clair and Hindar (2005), however, a brief overview of the different techniques is set out below.

Calcium carbonate can be applied directly into rivers by several different methods. The simplest is to place limestone sand or gravel in piles in the river channel (termed point application in this review) and the most sophisticated is to use machines which continuously apply calcium carbonate into the water (termed dosers). Dosers work by a variety of methods but in general all vary the dose applied with river flow rate; the most sophisticated of these continually monitor pH and adjust the dose, the simpler are flow driven. Complex dosers require the monitoring and dosing equipment to be continuously maintained and so can be expensive (Clair and Hindar 2005). Lake liming also involves the direct application of calcium carbonate into the water. Calcium carbonate is directly applied by boat, helicopter or tractor into the water or spread over a frozen lake in winter. The frequency of application depends on the retention time of the lake. In lakes with a long retention time one dose can increase the pH for several years (Donnelly et al. 2003).

Catchment liming differs from river and lake liming by applying calcium carbonate to the terrestrial environment rather than to the water body directly. It can be applied either to the entire catchment area or only the source areas, by hand, tractor, or helicopter. By applying calcium carbonate to the catchment the release of potentially toxic metal ions (e.g. Al^{3+}) from the soil can be decreased by increasing the pH of the water flowing through it (Clair and Hindar 2005). Catchment liming can also be expected to have a longer-term impact than individual direct applications to water, but all areas where calcium carbonate is applied are directly altered and it may negatively impact on the functioning and diversity of naturally acidic bogs and terrestrial plants (Donnelly et al. 2003). The wider impact of terrestrial liming will depend on the proportion of the catchment which is limed, the location of applications and the presence or absence of sensitive species in those locations.

With all liming methods the most commonly used liming material is limestone ground to a gravel or powder, although dolomite, $CaMgCO_3$, is also used within liming programs (Claire and Hindar 2005). The dose applied can vary greatly within each liming method and is generally calculated by modelling the neutralizing requirements of the catchment (Donnelly et al. 2003). The requirements will depend on the characteristics of the liming method chosen and the site being limed including: the type of application, desired increase in pH, water flow, size of the catchment, acidity of the catchment and the atmospheric inputs.

Liming, by all different methods, has been implemented in North America and many European countries. The largest liming programs have been undertaken in Norway and Sweden, and aim to restore the ecology of acidified rivers (Clair and Hindar 2005). Despite reductions to some liming operations in Scandinavia due to decreased acidic depositions (Barlaup 2004), liming still occurs widely in Europe. Liming is presently occurring both to protect rivers from acidification and to aid recovery as acidic depositions decrease. Within Europe, the Water Framework Directive requires member states to "protect, enhance and restore all bodies of surface water" to "good ecological status" (EU 2000), focusing governments attention on improving the ecological status of all water bodies. Within the assessment of the ecological status of water bodies set out by the directive the acidification status is included as one of the physico-chemical elements that supports the biological elements (EU 2000).

Liming is a potential option that may restore acidified waterways. However, liming only has an impact while it is in operation and hence would have to be applied as long as acid deposition outstrips the local soil neutralizing capacity (Donnelly et al. 2003). Additionally, it can be challenging to mitigate for episodic acidifications that can occur even once the overall level of acidification has decreased. Liming projects can be costly; Sweden has invested 3.8 billion SEK (approximately 0.4 billion Euros) on liming between 1983 and 2006 (Bostedt et al. 2010). Given the potentially high costs, it is important to ascertain whether liming will be an effective strategy or not. In some studies liming has increased targeted fish populations (Hesthagen and Larsen 2003, Clair and Hindar 2005) but liming has not always led to increases in fish numbers and the reported effect on invertebrates is variable (Clair and Hindar 2005).

There has previously been no systematic review of the evidence of effectiveness of liming to restore fish and invertebrate populations in rivers affected by anthropogenic acidification. Therefore this systematic review aims to find and summarise the best available evidence to address this deficiency. A linked review on the impact in lakes has also been produced, CEE 11-003.

2. Objectives

2.1. Primary objective

The primary objective is to answer the question:
What is the impact of 'liming' streams and rivers on the abundance and diversity of fish and invertebrate populations?

The question contains the following components:

Population: Fish and invertebrates within streams and rivers affected by anthropogenic acidification.

Intervention: Liming of streams and rivers, by any method.

Outcome: Change in abundance and diversity of fish and invertebrates.

Comparator: No intervention or before-after studies or both (before after control impact studies - BACI).

The presence of a comparator is important as it allows the assessment of the impact of liming relative to no liming. Without a comparator it is not possible to assess the impact of liming relative to no intervention.

3. Methods

3.1. Question formulation

This review is part of a collaboration between Environment Agency Wales (EAW) and the Centre for Evidence Based Conservation (CEBC). The review question was formulated following consultation with the EAW's policy group and EAW's staff at Cardiff and Swansea who formed the client group for the review. The review was conducted according to an a-priori peer-reviewed protocol (Mant et al. 2010)

3.2. Search strategy

3.2.1. Search terms

Search terms were chosen to capture as much of the relevant information as possible. They are separated into those relevant to the subject populations in terms of geographical area (i.e. streams/rivers), the subject populations in terms of the biota and the intervention. No terms for the outcome (i.e. change, abundance, diversity) were included in the search as these terms were not always included in the titles or abstracts of relevant studies. For each category included, different variations of the terms were used in order not to miss relevant papers, for example, the subject (ecological) terms not only included river and stream, the main subject of the review, but also other words for these terms such as brook and creek. "*" denotes wildcard.

Population - ecological: stream, river, catchment, brook, creek, burn, fluvial, source area, gravel.

Population - biota: fish* (includes fishes, fishery etc.), salmo*, trout, macroinvertebrate*, invert*, macrofauna, meiofauna, insect*, ephemeroptera, plecoptera, trichoptera, mollus*, crustacea*, microcrustacea*, bivalve*, gastropod, zooplankton, coleopteran, chironomid.

Intervention: liming, lime*, chalk*, calcium carbonate, dolomite.

Where possible the search contained all of the terms from each category. Terms within categories were linked with the Boolean operator 'OR' and terms between categories were linked with the Boolean operator 'AND'. Search engines vary in how many combinations they allow but for all databases the terms were combined in the most efficient way possible (see Appendix A for the details of the search strings used for each database). For example, Web of Science allowed combinations of all terms in one search.

3.2.2 Databases

The search covered the following databases which cover the breadth and depth of available literature on the topic:

1) ISI Web of Knowledge
2) Science Direct
3) Directory of Open Access Journals
4) Copac
5) Index to Theses Online
6) CAB Abstracts
7) ConservationEvidence.com
8) CSA Illumina
9) Agricola
10) Scopus

No time, language or document type restrictions were applied. The use of English search terms could have biased the findings against papers in other languages. However, this will have been reduced due to non-English language papers often still providing an English abstract and/or title. Additionally, the bias was mitigated as much as possible by searching the databases of specific organisations (see section 3.1.4).

3.1.3 Websites

An Internet search was also performed using meta-search engines and recommended sites:

http://www.alltheweb.com http://www.dogpile.com http://www.google.com
http://scholar.google.com http://www.Scirus http://data.esa.org/

In each case, the first 50 hits were examined for appropriate data, as recommended by the CEE review guidelines (CEE, 2010).

3.1.4 Specialist sources

Swedish and Norwegian environmental agencies were contacted and websites searched.

Websites of relevant specialist organisations, listed below, were also searched. Each website was searched using the website's own search facilities where possible; where no search options were available, websites were searched manually. In all cases if the website contained sections for reports and publications, this part of the website was searched in detail. Website search engines, where present, generally only accepted simple search terms. Hence, not all search terms for the main database search were used. The searches were restricted to liming, lime* and the term for liming in the native language of the website (i.e. kalking for Norwegian sites and kalkning for Swedish sites).

Anglers Trust
Alterra
British Ecological Society
Centre for Ecology and Hydrology
Countryside Council for Wales
Department for the Environment, Food and Rural Affairs
Dŵr Cymru / Welsh Water
Environment Agency
Environment Canada
Environmental Protection Agency
Environmental Protection Agency Ireland
EHS –Northern Ireland Environment Agency
European Commission Joint Research Centre
European Environment Agency
Finnish Game and Fisheries Research Institute
Fisheries research service
IKEU, Swedish Environment Protection Agency's national liming monitoring program
International Union for Conservation of Nature
Joint Nature Conservation Committee
Macaulay Land Use Research Institute
National Parks Natural England
Natural Resources Canada
Norwegian Directorate for Nature Management
Norwegian Institute for Nature Research (NINA)
Norwegian Institute for Water Research
Research Councils UK
Severn Trent Water
Salmon and Trout Association
Scottish Agricultural College
Scottish Executive
Scottish Environment Protection Agency
Scottish Natural Heritage
Society for Ecological Restoration
United States Environment Protection Agency
United Utilities
Welsh Assembly Government
Yorkshire Water

Bibliographies of included material were searched for relevant references.

3.3 Study inclusion criteria

All articles found by the above search strategy were first sorted for duplicates. The articles were then subject to a three stage process to identify the most relevant articles for the review question. The aim of this process was to systematically remove articles that were not relevant or did not contain relevant information or data. At each stage, if there was insufficient information to exclude an article it was retained until the next stage.

In the first instance, the titles of the articles were assessed using the inclusion criteria set out below, in order to remove spurious citations. Articles remaining after this filter were filtered on viewing the abstract and then the full text.

In order to pass each stage each article had to meet the following criteria:

• Relevant population (s): the biota in any stream, river, or catchment that is or has been affected by the effects of anthropogenic acidification[1].

• Types of intervention: Addition of calcium carbonate (or dolomite) to ameliorate the effects of acidification in streams, rivers, and catchments. Methodologies include hydrological source liming, point source liming, doser liming, stream liming, river liming, and catchment liming. No particular method of liming was excluded.

• Types of comparator: No intervention or before-after comparisons or both (Before- after control impact studies – BACI).

• Types of outcome: Change in abundance, density, diversity or richness of fish or invertebrates.

• Types of study: Any primary study comparing limed and un-limed subjects whose outcomes fit the above. Review articles will not normally contain primary data but were searched for the primary studies they include. No geographic restriction was applied to this review.

In order to assess and limit the effects of between-reviewer differences in determining relevance, two reviewers applied the inclusion criteria to 200 articles (over 20%) at the start of the abstract filtering stage. A kappa statistic (Edwards et al. 2002) of 0.6 was calculated, which measures the level of agreement between reviewers and represents a reasonable level of agreement.

3.4. Study quality assessment

All articles accepted at full text were critically appraised according to their study design and quality. Well-conducted studies of high quality have less potential for bias than their poorer counterparts.

Firstly the studies were categorised according to the study design into:

i) Before after control impact (BACI) studies
ii) Control impact (CI) studies (using an upstream section of the same river)

---

[1] Identification of the level of anthropogenic impact on the pH of a watercourse can be challenging, therefore, relevant studies were those where liming was carried out to mitigate assumed human induced acidification. Liming to mitigate acidification caused by acid mine drainage was not included.

iii) Control impact (CI) studies (using limed control streams)
iv) Before After (BA) studies

BACI studies are less prone to hidden bias than BA or CI studies. In CI studies the baseline of the treated and untreated rivers are not recorded, hence, differences observed after treatment may be due to differences before treatment rather than the treatment itself. Random assignment to treatment groups can ensure that variability in baselines would be equal in each groups, hence how study sites were selected and assigned was recorded. In BA studies the situation after treatment is compared to the situation before treatment, there are no untreated controls; differences may be due to factors that have changed during the course of the observations other than the intervention. Hence, the location of the study and any changes that occurred other than the intervention were also recorded during the study quality assessment. Additionally, the reliability of studies depends on the level of replication within the study so the level of replication was also recorded, both within a river and the number of rivers sampled.

3.5. Data extraction

Along with the information on the study design and quality, the outcomes of the studies were recorded. Data on changes in invertebrate and fish abundance and diversity (or species richness) were extracted; as many data as possible were extracted from each paper including the means, the number of observations contributing to the mean and the variances. From the presented data, effect sizes were calculated wherever possible. The effect size is a measure of the magnitude of the impact of the intervention (liming). There are several different types of effect size that can be calculated but for this study the log ratio has been used. The log ratio was chosen as being more appropriate compared to other common effect size measures; mean difference and standardized mean difference. The mean difference was not used as all studies would have been required to have the same measurements (units) and this was not always the case. Also, the mean difference gives the absolute increase in number, however, an increase of one fish per 100 metre squared would be more significant if there were only a few before hand rather than if there were hundreds. The log ratio gives the relative increase taking into account the starting sizes and is dimensionless. The standardised mean difference is also dimensionless as the mean difference is divided by the standard deviation of the mean. However, small sample sizes decrease the accuracy of the standard deviation estimates and many of the studies had small sample sizes. Using variances from small sample sizes, and which may have been calculated across different scales of variation, would increase the potential inaccuracies in the effect size estimates, hence, the standardised mean difference was not used.

The log ratio was calculated using the natural log of the ratio of treatment to control. In terms of the different study designs encountered the ratio is calculated as:

For BA studies; the ratio =A/B
For CI studies; the ratio = I/C
For BACI studies; the ratio = $(I_A/I_B)/(C_A/C_B)$, which is the same as $(I_A/C_A)/(I_B/C_B)$

In the main results section the results are presented as the log ratio data calculated from the analysis. Thus, a decrease is represented by a log ratio of less than zero, a minus number), and an increase is represented by log ratio of greater than zero. It is possible to calculate the response ratio by back transforming log response ratio, which has been done in the conclusions, for ease of interpretation; hence a ratio of less than 1 represents a decrease and a ratio of greater than 1 represents an increase (rather than greater than zero as for the logged ratio).

In order to provide the most accurate estimate of the true effect size for paired data, the ratio was calculated for each pairing separately and then averaged rather than calculating the ratio on the average for each grouping. Details of how the effect size was calculated for each study can be found in Appendix B.

The error in the effect size was calculated either by using the formula (4.23, p31) from Bornstein et al. (2009) or if data were paired by taking the standard error of the variance in ratios over the ratio for each pairing.

3.6 Data synthesis

Meta-analyses were carried out on the extracted effect sizes using the R package 'metafor' (Viechtbauer 2010). Random effects meta-analyses, weighting by inverse variance, were carried out using the DerSimonian-Laird estimator method. Random effects models were used rather than fixed effects models as there are likely to be differences in the effect measured in different studies that are not just due to chance. The weighted mean effects, confidence in the mean effect and prediction interval were calculated for each of the variables analysed: fish abundance, fish species richness, invertebrate abundance, number of invertebrate taxa, acid sensitive invertebrate abundance and number of acid sensitive invertebrate taxa. The mean effect is the mean of the effect sizes from each included study weighted by the inverse of the variance in the effect size. The confidence interval for the mean effect is the interval in which we are 95% confident that the mean effect occurs. As there will be several factors that vary between studies not all liming operations are likely to experience the mean effect; the "true effect" will vary between studies. Hence, the prediction interval is also calculated: the interval giving the distribution of effects across studies/liming operations/rivers, within which 95% of true effects are predicted to occur. Where there were sufficient data the impacts of potential effect modifiers were tested; type of study, type of liming, presence of stocking and the mean length of time over which calcium carbonate was applied. Fish abundance provided the largest dataset and hence the greatest opportunity for the analysis of potential effect modifiers. As the effect of the liming operations was calculated from all available data, including in many cases multiple years during liming, the length of the liming operation in a study was calculated as the number of years liming had been occurring averaged over all of the samples taken in the study (i.e. if a study presented data for each of the first 10 years of continuous liming the average length of liming over those 10 samples would be five years). Catchment liming operations are often one-off applications that are expected to impact over a long timeframe, however, there was only one study (out of 33) within the analysis of length of liming operation that only used catchment liming so despite it being treated as a short term application it is unlikely to have significantly impacted on the results.

4. Results

4.1 Studies found

The search was carried out between the 20/7/2010 and 4/8/2010 and identified 3641 potentially relevant articles. Many were actually paleontological articles and 937 articles were retained after the title inclusion stage (Figure 1). After the abstract inclusion stage 358 articles were included with an additional 35 articles from organisational searches and bibliographies found in the search. At the full text assessment a total of 72 articles were found to be relevant. Many studies were not included in the review because they covered biota in lakes but not streams, were about acidification due to mining, or the studies only looked at the impact of liming on the physiology of organisms, not the abundance or diversity (species richness) of species within the ecosystem (see Appendix C for the full list of studies excluded the full text stage). One study looked at the impact of liming to mitigate acidic soils in the tropics but this was excluded from the review as being a different system, in all other papers the authors stated that the acidification was assumed to be caused by acid rain or did not explicitly state the cause but did not mention acidification due to acidic (sulphate) soils. Additionally to the main list of included studies, the main dataset of the monitoring program of the Swedish national liming project and the latest report from the Norwegian national liming program were obtained.

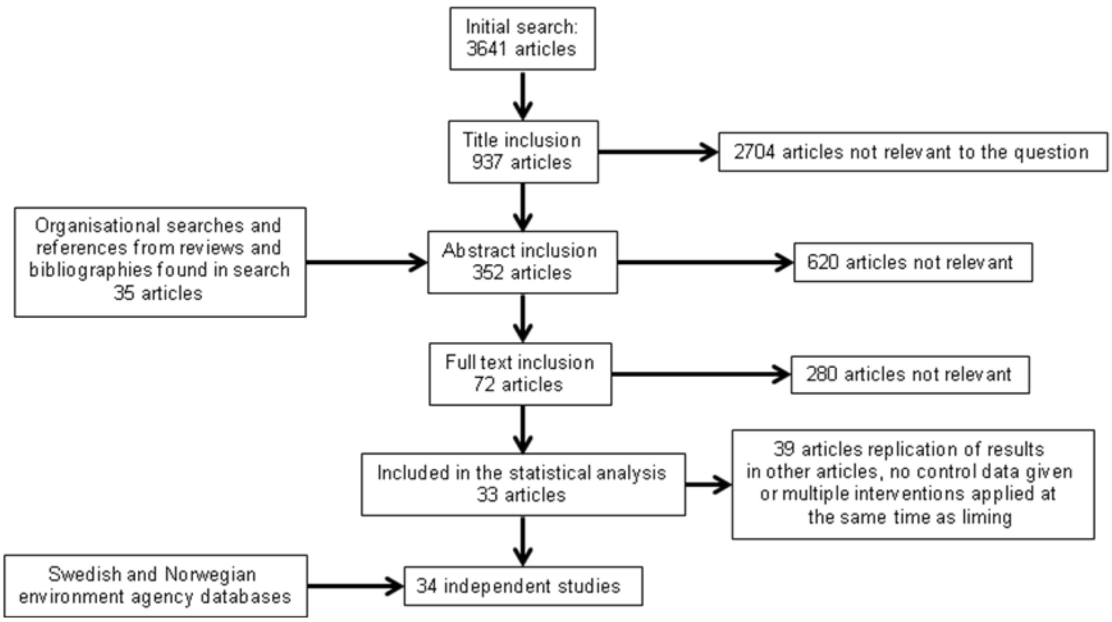

Figure 1: Flowchart of process for identifying the articles and studies they contained which were relevant to the review question

The articles relevant to the review were then assessed for inclusion in the analysis. In order to avoid pseudo-replication, 33 articles were excluded due to overlap in the reported data (i.e. where the same study was reported in different formats); for each river studied the impact of liming was only recorded once for each outcome measure (i.e. fish species richness, invertebrate abundance etc.). Additionally, a further 6 studies (and one river from the Norwegian survey) were excluded as relevant data could not be extracted from the article (i.e. due to the study being poorly reported, a lack of appropriate controls or multiple interventions being applied at the same time; for details see Appendix B). This left 33 articles plus the Norwegian survey and the Swedish dataset. In total these 33 articles and 2 datasets covered 45 studies (19 of which were rivers in the Norwegian survey and one of which was the main Swedish study that covered 18 limed rivers and 8 acid control rivers; details of all studies are presented as a narrative table in Appendix B). Of the main 26 studies not in the Norwegian survey the majority (15) were on single rivers or streams and only three (all from Sweden) were on 10 or more rivers or streams (Figure 2).

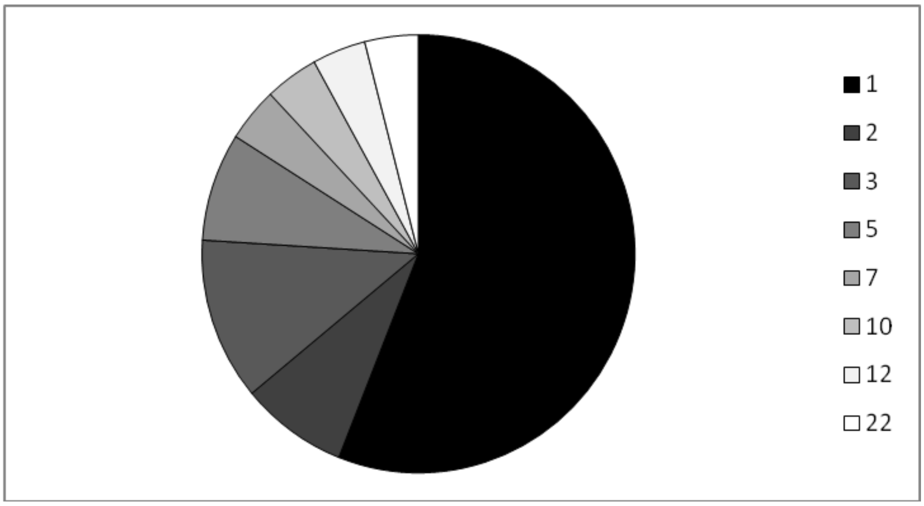

Figure 2: Pie chart of the proportion of studies that covered different numbers of rivers; black represents study on one river, increasing lightness of colour shows increasing number of rivers covered where white represents 22 rivers covered in one study.

The selected studies were not evenly distributed across the world. Studies came from Norway (19), Sweden (7, including several large studies), USA (11), UK (5), Canada (3) and France (1). The problems of acidification are also not evenly distributed across the globe. Scandinavia has experienced the largest acidification problems in Europe due to a combination of acid sensitive water bodies and being located downwind of major sources of atmospheric pollution. Hence, the uneven distribution of studies can be explained by uneven distributions in the problem rather than bias in the study collection.

The studies were published from 1983 to 2009 (Table 1). The largest number of studies was published in 1995-1999 (excluding the latest version of the Norwegian survey that covered 19 studies, published in 2009). As the majority of studies were published after 1995 they are likely to include data from the period of time when acid depositions have been decreasing.

Table 1: Year of publication of the studies

Year of publication	Number of studies
1983-1989	3
1990-1994	4
1995-1999	10
2000-2004	4
2005-2008	4
2009 (Studies in the Norwegian survey)	19
2010 (Swedish database)	1

Information on both fish and invertebrates were included in the studies. In total there were 32 studies on fish and 29 studies on invertebrates (several studies included both fish and invertebrates so the total is greater than the total number of studies). The included studies also covered the wide variety of liming methods in operation including lake liming, catchment liming, point applications into river and continuous dosing into rivers (Table 2).

Table 2: Liming methods used within the studies found

Liming method	Number of studies
Catchment liming	6
Lake liming	3
Doser	9
Point application into a river	9
Doser and lake liming	13
Doser, lake and catchment	1
Method varied between rivers in study	3
Method unclear	2

4.2. Study quality assessment

The majority of the included studies presented sufficient data for effect sizes to be calculated (only six studies were not included in the quantitative analysis due to a lack of appropriate data, see Appendix B). There was a variety of study designs present and even within one study the design often varied between the fish and invertebrate data. However, studies broadly fell into three types of design; before-after studies (BA), control-impact studies (CI) and before-after-control impact studies (BACI). For BA studies, as the studies were taken in the natural environment (rather than a laboratory), it is not possible to control for potentially confounding factors due to variations in other factors during the course of the experiment. Acid deposition has been decreasing over the last two decades (Evans et al. 2001) and this decrease in the cause of acidification has the potential to introduce a major bias and confounding factor to BA experimental designs. The importance of this bias will depend on the timescale of the experiments; over a short time scale acid deposition is less likely to have changed noticeably. Only three of the 21 BA fish studies took measurements for less than an average of three years after the start of liming. Over half of the fish BA studies took measurements for an average of six or more years from the start of liming (see Appendix B for details).

Control impact studies can have potential problems if there are underlying differences between the control and treated rivers. Randomizing the assignment of each river to the control or treatment group decreases the chances of systematic differences between the control and treatment rivers. However, in none of the studies with control sites was it clear that there had been random allocation of groups. The controls were either upstream sections of the limed river or unlimed control rivers. Where the controls were upstream sections there was clearly not random allocation as the controls were selected due to the nature of the site (i.e. being upstream of the liming point). Rivers and streams naturally vary along their length and hence differences in fish and invertebrate populations can be expected along the length of a stream. The differences between upstream and limed sections will depend on the proximity to the limed region and this was variable. In the main Norwegian survey (seven rivers) the control sites were often in unlimed tributaries which may have been different. In smaller scale experimental studies the controls sites were generally closer (i.e. Little Stoney Creek; Downey et al. 1994 and Liscome; Watt et al. 1983).

Studies where the controls were independent unlimed rivers do not suffer the same problem but still none of the studies clearly showed randomized allocations. In half of the studies (four out of eight) there was mention of some form of 'matching' or proximity of the control streams but it was not clear how this was done or how closely matched the sites were. Also, in several of the studies the liming operations were already in place and the limed sites were selected by those responsible for the operations and not the researchers carrying out the studies. It is unclear how liming operators chose which rivers to dose. Limed rivers may have been the more 'ecologically' important ones with higher species abundance in the past because operators and managers of liming projects will want their intervention to have the largest possible effect. Additionally, economic factors may have influenced the decision about which rivers to dose driven by the desire to restore high economic value game fisheries or simply liming the easier to access sites. Thus, differences observed in the control impact studies may be due to underlying differences in the rivers chosen to be limed.

The use of before-after-control-impact (BACI) study designs can control for some of the bias described above. The use of a control allows changes not due to liming to be monitored and the use of before data can highlight underlying differences in the control and treated sites. However, if differences are present in the baseline data, the control and limed sites may not respond in the same way; five of the BACI designed experiments showed differences in baseline measurements for at least one of the measured outcomes.

For the fish studies, the majority (20 out of 31) were BA studies (Table 3). Only seven of the fish studies were BACI studies. The majority of the invertebrate studies were of CI design (16 out of 29). A higher proportion were of BACI design than for fish but it was still only ten studies. Details of each of the studies designs and potential confounding factors are available in Appendix B.

Table 3: The number of studies measuring fish and invertebrates which had each study design.

Study design	Fish	Invertebrates (all invertebrates and acid sensitive invertebrates
BACI	7 (23%)	10 (34%)
CI	4 (13%)	16 (55%)
BA	20(64%)	3 (10%)

In addition to the factors described above, the presence or absence of stocking can introduce another important confounding factor. Many of the studies within the Norwegian survey reported the use of stocking to aid the recovery of fish stocks from acidification. The stocking of fish is likely to alter fish abundance and if stocking starts after liming, this will be an important confounding factor. Most studies outside Norway did not mention stocking, although it may still have occurred and not been reported.

The reliability of a study is also determined by the level of replication employed. Most of the studies were on only one river (Figure 2), hence they provide estimates of the impact of liming on that river but it is hard to generalise conclusions across to all rivers from one study. Single river studies will not provide good estimates of the mean impact on a population of rivers.

The reliability of the estimate of liming impact on a single river will depend on the level of replication within the river, the area sampled within each replicate and the variability within the river. If a river is very variable along its length then one site is unlikely to be representative of the number of taxa or abundance everywhere. In the Swedish dataset, very variable numbers of species were found at different sites along a single river and only a few sites were sampled per river. Hence, not all species present in the river will necessarily have been sampled by the surveys. Also, differences between sites may have been exacerbated by nearly 10 fold variation in the area surveyed at each site. Other studies had differences in their level of sampling in the treatment and control groups (see Appendix B). Rivers are not uniform along their length but if a large number of samples are collected, there is a greater likelihood that the full extent of the variation will be sampled. The studies varied widely in the number of samples taken in both space and time within one river. The sampling period varied from one sampling occasion to 22 years of yearly sampling and the number of sites varied from one to 20 (see Appendix B for details). The studies with the longest time series were not necessarily the ones with the largest number of sites surveyed; also the studies with the most samples in time and space were often not the studies with BACI design.

4.3 Quantitative synthesis/Meta-analysis

Fish abundance

The effect of liming on fish abundance varied significantly among the included studies (Figure 3, random effects meta-analysis on log ratios, Q=363, df=33, p<0.001). The observed effect varied from a negative impact in some rivers to a positive impact in others. Overall, there was a mean positive effect (mean effect (log ratio) =0.53, SE=0.14, z=3.63, p=0.003), however, the prediction interval, the interval in which 95% of true effects are predicted to occur, overlapped zero (log response ratio of -0.4 to 1.5). Hence, the meta-analysis suggests that if liming is adopted as a national policy the average impact would be to increase fish abundance but, in any one river there is an 18% possibility of the fish abundance decreasing after liming (assuming a normal distribution of true effects of the log ratio). A variety of abundance measures was used in different studies including; population estimates, density estimates of number of adult fish per 100 m^2 or number of fry and par per 100 m^2, fish biomass in kg/ha and fish biomass in total kg caught per year. The baseline abundance estimates for the three most common measures, number of fry and par per 100 m^2 (n=17), number of adult fish per 100 m^2 (n=4) and total kg caught per year (n=5) are respectively; 36 per 100 m^2 (SD 25), 5.1 per 100 m^2 (SD 6.1) and 186 kg/year (SD 91).

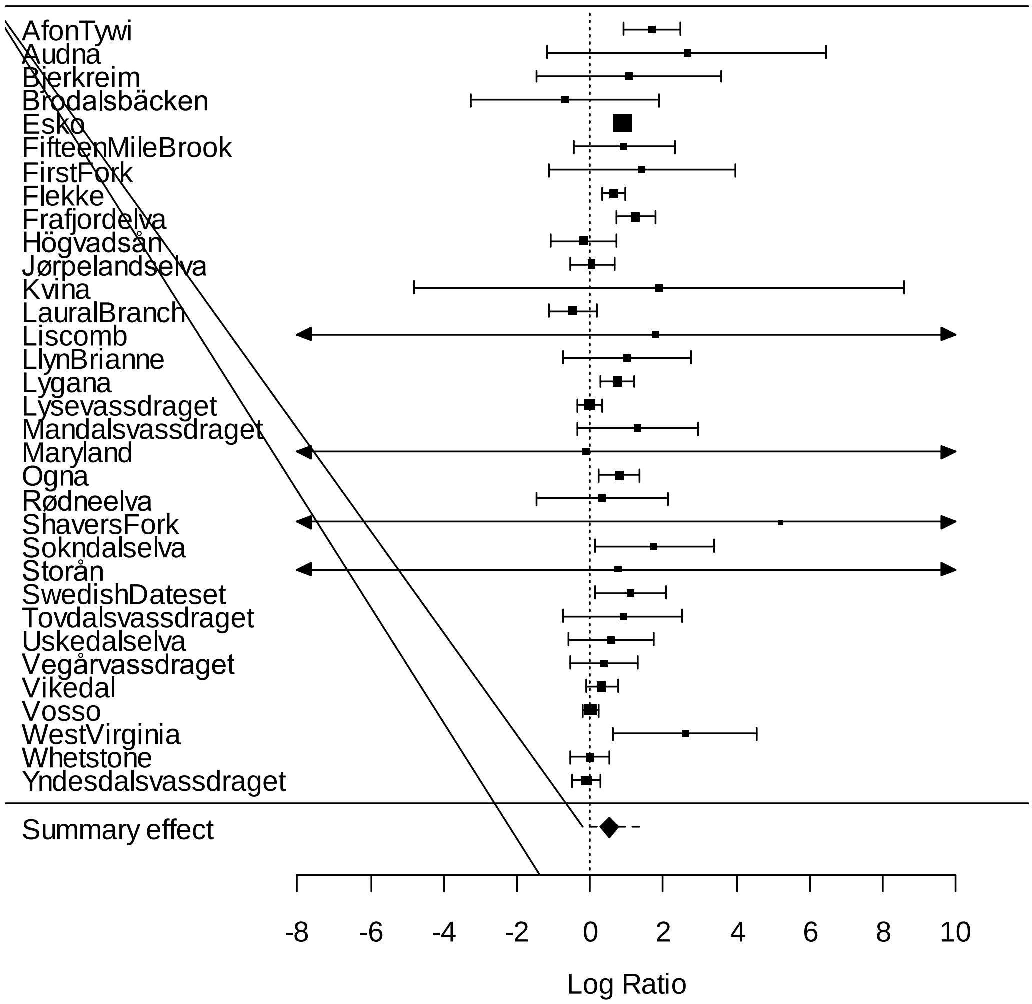

Figure 3: Forest plot of the fish abundance effect size for each of the studies. The squares are the effect size for each study, the error bars are the confidence interval for each study (arrows indicate a larger confidence interval (shavers CI -56 to 66) or where insufficient data were presented in the study to calculate a confidence interval (Liscomb, Maryland and Storan)). Summary effect: the diamond represents the weighted mean calculated from the random effects meta-analysis, the width of the diamond is proportional to the estimation in the error of the mean and the dotted line is the prediction interval were 95% of true effects are predicted to lie.

The finding of a significant mean effect shows that where liming has taken place the abundance of fish, on average has increased. However, the differences could be due to differences other than liming (see section 4.2). The chances of differences being due to factors other than liming are reduced within a BACI experimental design. Only a minority of the fish abundance studies were of a BACI design, however, there was not a significant difference in the effect size generated by different study designs (study type was not a significant effect modifier within the random effects meta-analysis, Qm=0.796, df=1, p=0.372). Hence, the study design does not appear to have caused the observed effects.

The variability in effect sizes between studies, heterogeneity, indicates that factors other than the liming intervention are influencing the effect of liming. If study design is not causing the variation, the differences could be caused by differences in the method of liming and characteristics of the sites. Liming method (river, catchment or lake) and the presence of fish stocking were not significant effect modifiers (Qm=0.559, df=1, p=0.455 and Qm=0.587, df=1, p=0.444 respectively) but the average length of time the river had been dosed was a significant modifier (Qm=4.239, df=1, p=0.039, Figure 4). As the effect of the liming operations was calculated from all available data, including in many cases multiple years after liming, the length of the liming operation in a study was calculated as the number of years liming had been occurring averaged over all of the samples taken in the study (i.e. if a study presented data each of the first 10 years of liming the average length of liming over those 10 samples would be five years).

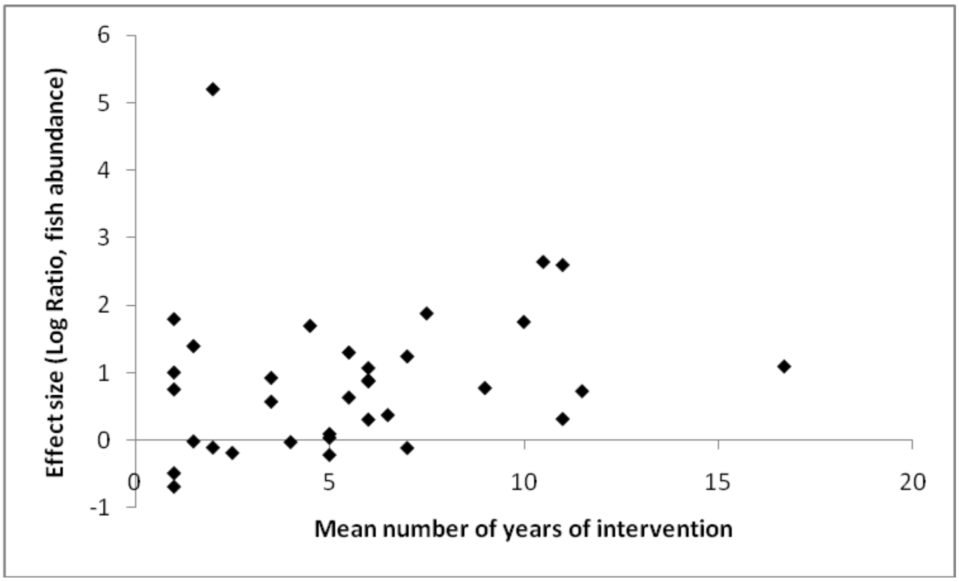

Figure 4: The relationship between duration of liming intervention and effect of liming on fish abundance.

As the number of treatment years increases the effect size generally increases (Figure 4). However, the study in which there was the largest effect size had only been limed for an average of two years (the Shaver Fork study). None of the rivers which had been limed for an average of more than 7.5 years showed a negative effect size. A confounding factor to this finding is that the scale of the liming operation was generally linked to the duration of the intervention. Small scale experimental set ups were often only operating for a few years whereas large-scale national operations, e.g. those in Norway and Sweden, were often in operation for many years. Additionally, the mean duration of application does not explain all of the variation, so other factors must be important.

Several studies presented the change in fish abundance in terms of particular species of fish rather than the overall change in the abundance of all fish. The most common fish presented were Atlantic salmon (*Salmo salar*) and brown trout (*Salmo trutta*). Where both salmon and trout abundances were recorded from the same river the effect sizes were different for the two species, showing that not all fish species react in the same way (Figure 5). In all except one of the 19 rivers where effect sizes for both could be calculated, salmon populations increased to a greater extent than trout. In that river, the river Vosso, the salmon density decreased whereas in all of the other rivers the salmon density increased after liming. There was significant heterogeneity in the salmon effect sizes (Q=114, df=18, $p<0.0001$), however, the mean effect was still a significant increase (mean effect (log ratio) =1.16, SE=0.38, z=3.02, p=0.003). There was also significant heterogeneity in the effect size of the trout (Q=302.3, df=18, $p<0.0001$), but the direction of mean effect was in the opposite direction from salmon, a small negative effect, although the confidence interval for the effect overlapped zero (mean effect (log ratio) =-0.15, SE=0.17). The trout population, when salmon were also present, increased in some rivers after liming but decreased in others and average effect over all rivers was unclear. In the three studies where brown trout were studied, and salmon were not present, the trout increased in abundance in each study (Swedish database log ratio = 0.79 (SE 0.5), Llyn Brianne log ratio = 1.0 (SE 0.9) and Whetstone log ratio =0.6 (SE 0.2)).

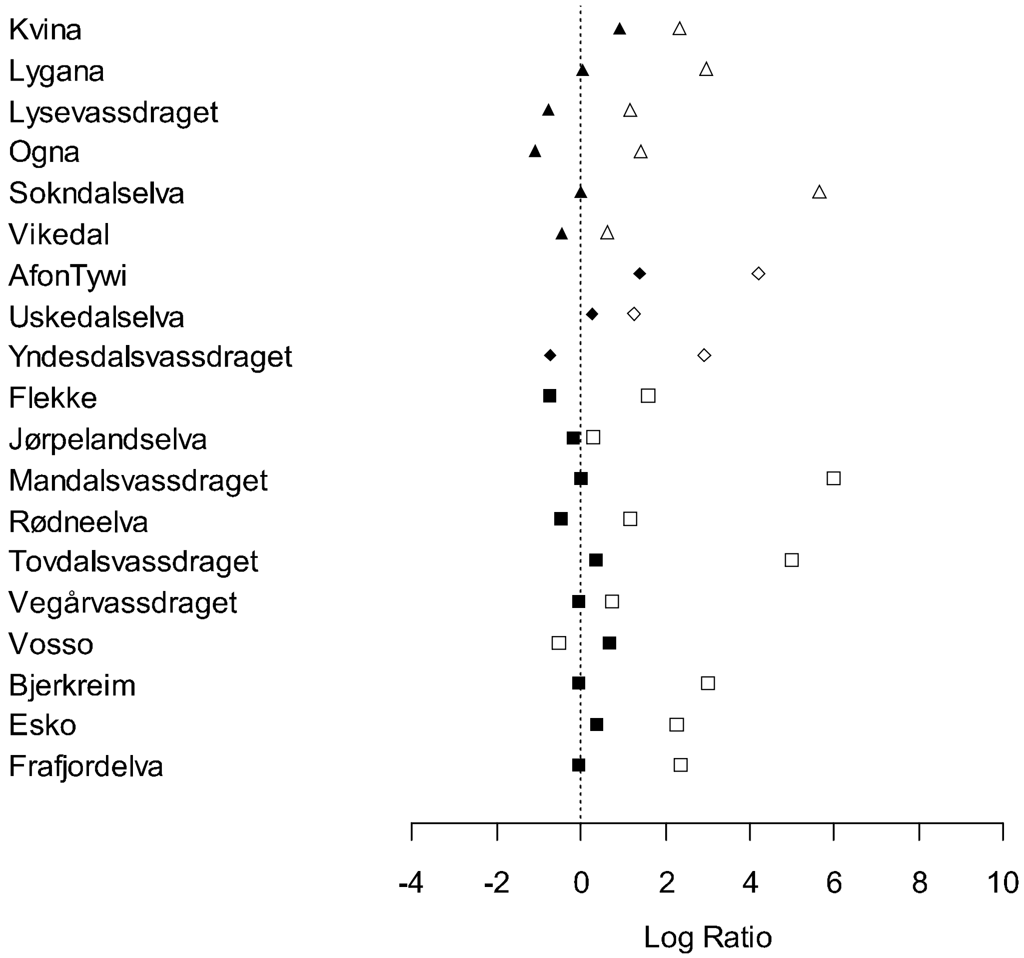

Figure 5: Graph of the effect sizes for salmon and trout abundance in studies where both were measured. The black symbols are the effect sizes for trout and the white symbols represent the salmon effect sizes. The graph is also divided on the presence of stocking; triangles represent where there was no stocking, diamonds where stocking was not mentioned in the articles and squares where stocking took place.

As salmon and trout respond differently where they are both present, it suggests that other fish species will also vary in their response due to both physiology and species interactions. There are very few data on non-salmonid fish, hence, this hypothesis cannot be tested and it is unclear how liming affects other fish species.

Fish Diversity

The changes in fish abundance can lead to changes in diversity; however, far fewer studies measured fish diversity than fish abundance. The diversity of fish was normally recorded as the number of fish species present rather than a diversity index. As with fish abundance, the effect of liming on fish species richness varied significantly between studies ($Q=44.2$, $df=6$, $p<0.0001$, Figure 6). The small number of studies and inability to calculate the error in the observed effect for two of them (Degerman and Maryland) meant that the confidence interval for the mean effect was large and overlapped zero (mean effect (log ratio) =0.47, SE=0.28, z=1.68, p=0.093). There is no clear evidence that the mean increase in abundance of fish after liming is likely to be accompanied by a mean increase in the number of fish species.

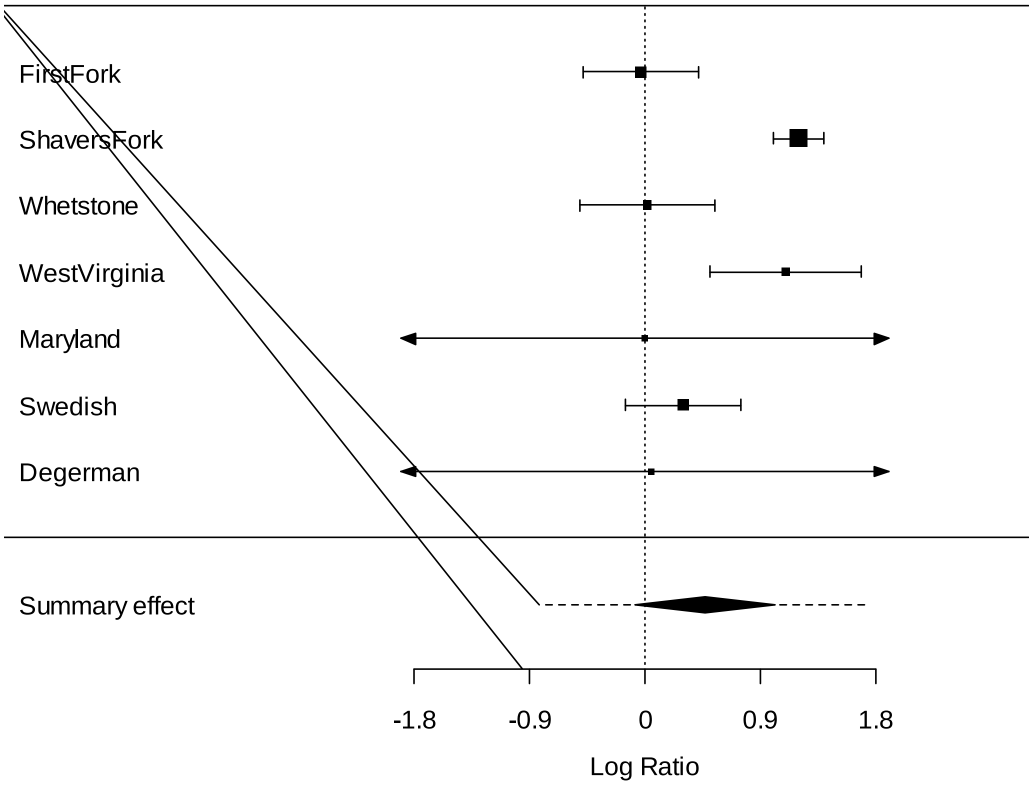

Figure 6: Forest plot of fish species richness effect size for each of the studies. The squares are the effect size for each study, the error bars are the confidence interval for each study (arrows indicate where insufficient data were presented in the study to calculate a confidence interval). Summary effect: the diamond represents the weighted mean calculated from the random effects meta-analysis, the width of the diamond is proportional to the estimation in the error of the mean and the dotted line is the prediction interval were 95% of true effects are predicted to lie.

Invertebrates

Liming will not only impact on the commercially important fish populations but also on the wider ecosystem. Invertebrates within streams are good indicators of ecosystem quality and are important within the food chain, including for fish. The impact of liming on invertebrate abundance varied significantly between studies (Q=117, df=12, p<0.0001, Figure 7). The mean effect was no change in abundance (mean effect (log ratio) =0.01, se=0.12, z=0.07, p=0.944) and the prediction interval (the interval within which 95% of true effects are predicted to occur) ranged from -0.90 to 0.92. Hence, it is predicted that in some rivers invertebrate abundance will decrease after liming and in others increase. The liming method is not a good predictor of this variation (Qm=0.05, df=1, p=0.82). However, there is a non-significant trend for the type of study (Qm= 3.77, df=1, p= 0.052) and the mean effect size for the BACI design studies is negative, whereas the mean for the BA studies is positive. This potential difference between the studies with more or less risk of bias suggests that the risk of bias within the BA studies may be affecting the results. If the BACI studies are taken alone then the mean effect is a significant negative effect (mean effect (log ratio) = -0.24, se=0.12, z= -1.99, p= 0.047), although there is still significant variability between the studies (Q= 27.88, df = 5, p<0.0001). Hence, invertebrate abundance may on average go down after liming, although more studies of a BACI design are needed to be confident in this assessment.

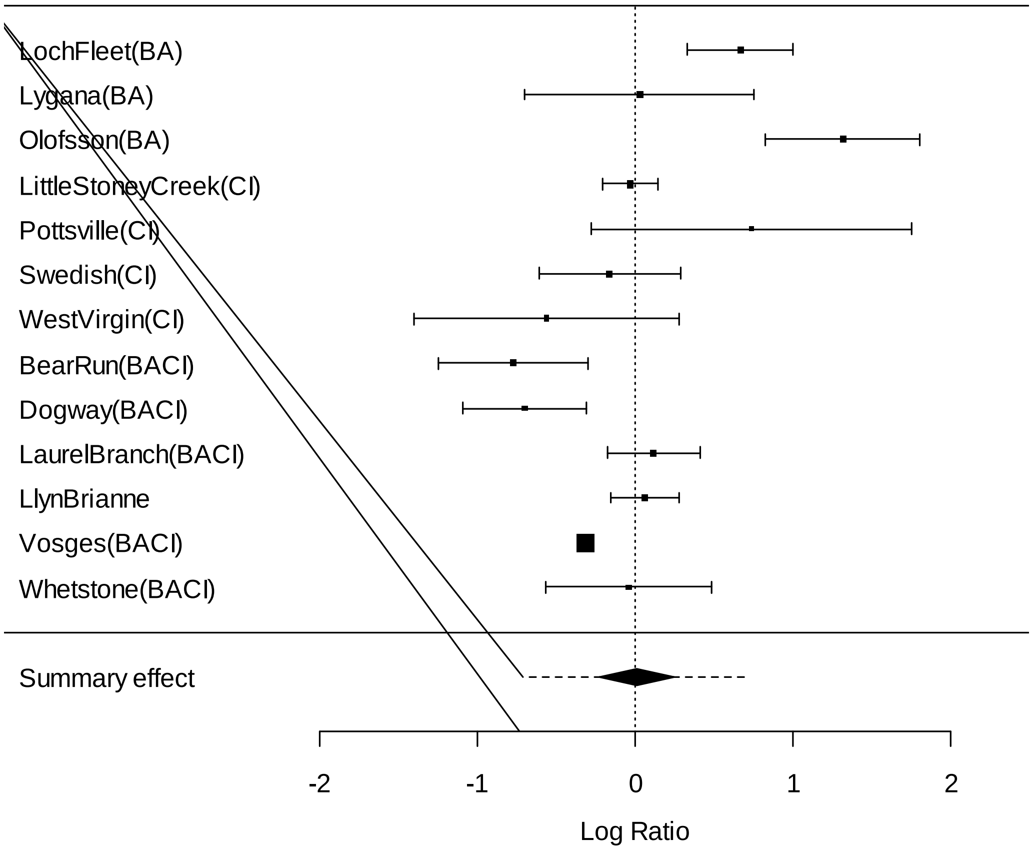

Figure 7: Forest plot of invertebrate abundance. The experimental design for each study is given in brackets after the study name. The squares are the effect size for each study, the error bars are the confidence interval for each study. Summary effect: the diamond represents the weighted mean calculated from the random effects meta-analysis, the width of the diamond is proportional to the estimation in the error of the mean and the dotted line is the prediction interval were 95% of true effects are predicted to lie.

The smaller number of invertebrate studies, and large range in different types of liming and combinations of types of liming meant that the importance of liming type for explaining the differences could not be investigated.

The impact of liming on the diversity of invertebrates was most commonly studied by assessing the changes in the number of invertebrate taxa rather than a diversity index. The definition of taxa, level of identification of organisms, varied between studies but was usually given as the lowest identifiable taxonomic grouping (see appendix B for details on each study). The number of taxa varied significantly between studies (Q=42.8, df=15, p=0.0002, Figure 8). The mean impact was a significant increase (mean=0.16, SE 0.06, z=2.48, p=0.013), however, the prediction interval overlaps zero; some of the true effects are predicted to be negative. The type of study and liming method are not good predictors of this variation. However, within this group of studies the type of study (i.e. CI, BA or BACI) was not a good indicator of the study quality; although BACI studies generally use a high quality method, not all those used here were well designed: in three cases (Esk, Vikedal and Herrmann), no error could be calculated due to the lack of replication or a lack of reporting despite employing a BACI design.

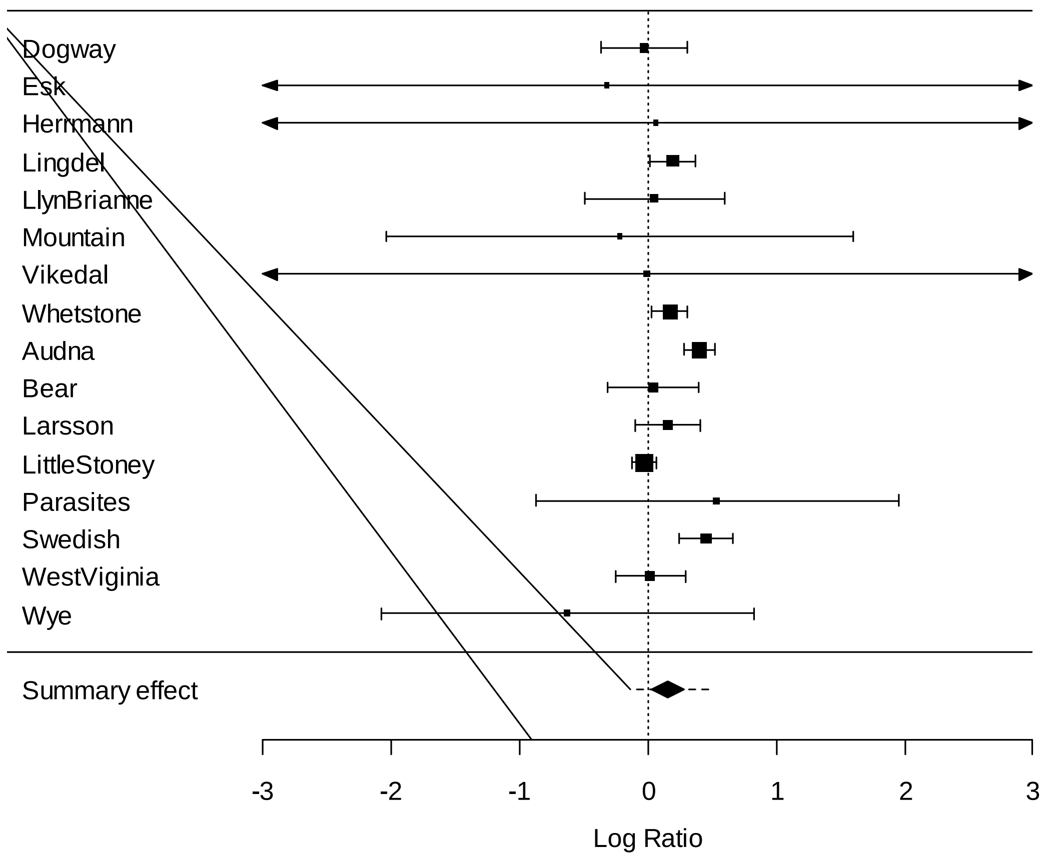

Figure 8: Forest plot of the effect size for the number of invertebrate taxa for each of the studies. The squares are the effect size for each study, the error bars are the confidence interval for each study (arrows indicate where insufficient data were presented in the study to calculate a confidence interval). Summary effect: the diamond represents the weighted mean calculated from the random effects meta-analysis, the width of the diamond is proportional to the estimation in the error of the mean and the dotted line is the prediction interval were 95% of true effects are predicted to lie.

Acid sensitive invertebrates

Liming is directed at particularly acid sensitive ecosystems or parts of ecosystems. Hence, the impact on acid sensitive invertebrates has been assessed separately from the response for all invertebrates. Unlike for all invertebrates, the impact of liming on acid sensitive invertebrate abundance did not vary significantly between studies (Q=10.4, df=5, p=0.06, Figure 9). However, the level of heterogeneity was still relatively high (I^2, the percentage of total variability due to heterogeneity = 52%). All of the measured effect sizes were positive and the mean effect was a significant positive effect (Mean= 0.68, SE=0.29, z=2.56, p=0.018), however, due to the high variability within studies and the small number of studies, the prediction interval overlaps zero.

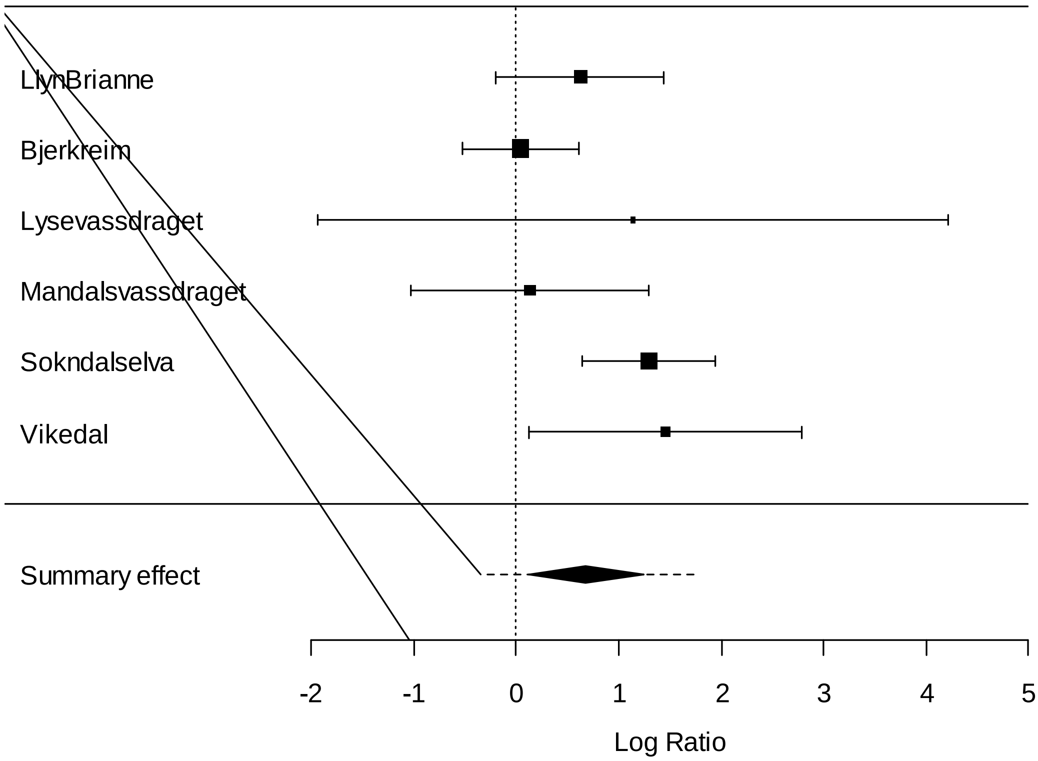

Figure 9: Forest plot of acid sensitive invertebrate abundance effect size for each of the studies. The squares are the effect size for each study, the error bars are the confidence interval for each study. Summary effect: the diamond represents the weighted mean calculated from the random effects meta-analysis, the width of the diamond is proportional to the estimation in the error of the mean and the dotted line is the prediction interval were 95% of true effects are predicted to lie.

The effect of liming on acid sensitive invertebrate diversity was also reported as the number of acid sensitive taxa present. The number of taxa did not vary significantly between studies ($Q= 0.52$, $df = 4$, $p= 0.97$, Figure 10) and unlike for abundance none of the variability was due to heterogeneity of studies ($I^2=0\%$). There was a significant increase in the number of taxa after liming (mean log ratio= 0.95, SE=0.23, z=4.16, $p<0.0001$) and the mean effect was larger than the mean effect for all invertebrates (0.95 compared to 0.16). Also, the two studies excluded from the analysis because log ratio effect sizes could not be calculated, strengthen this result. In both studies (West Virginia and Dogway) the number of acid sensitive invertebrate taxa was higher in the limed sites. In the West Virginia study there were no acid sensitive invertebrates in the control and in the Dogway study, the control increased from zero acid sensitive invertebrates, hence log ratios could not be calculated (see Appendix B). The confidence in the mean calculated in the meta-analysis is driven largely by the Audna study which has the smallest variance. If it is removed from the analysis, the mean effect is no longer significantly different from zero. The Audna study covered only one river, whereas the study where the confidence interval could not be calculated, thus with the smallest weighting, covered the largest number of rivers (five) and had a negative effect size.

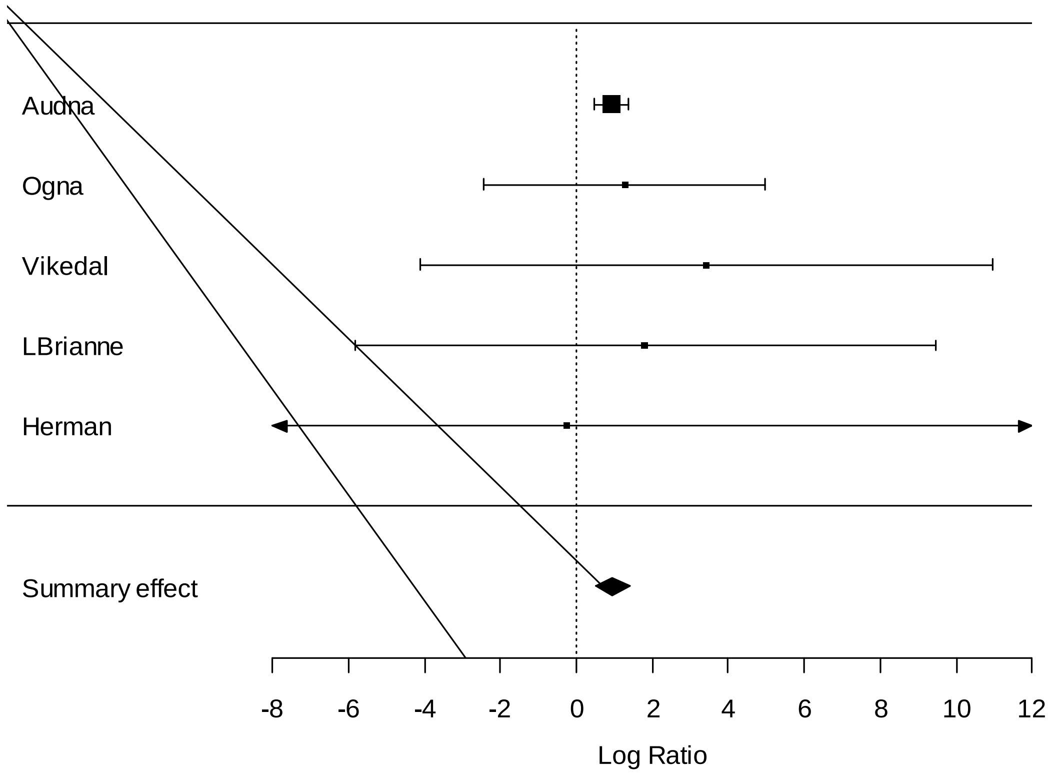

Figure 10: Forest plot of the effect size for the number of acid sensitive invertebrate taxa for each of the studies. The squares are the effect size for each study, the error bars are the confidence interval for each study (arrows indicate where insufficient data were presented in the study to calculate a confidence interval). Summary effect: the diamond represents the weighted mean calculated from the random effects meta-analysis and the width of the diamond is proportional to the estimation in the error of the mean (unlike in the other studies there is no dotted line from the diamond as the heterogeneity was calculated as zero).

5. Discussion

5.1 Evidence of effectiveness

In terms of fish abundance liming does appear to have an effect; on average mean fish abundance is greater in limed rivers. The size of this effect varies depending on the duration of the intervention, with longer interventions producing a larger effect, although there is still large variation in effect within this generalisation. As many of the studies are observational before after or control impact studies rather than large-scale randomised experiments, the correlations observed do not prove causation. The significant changes occurring over time may not be simply due to liming. Also the abundance of fish did not increase after liming in all rivers; in some rivers the numbers of fish actually decreased after liming (9 out of 33 studies had negative effect sizes for fish abundance). In some instances liming itself may cause negative impacts on fish directly, for example possibly due to changing substrate or causing toxic boundary areas between limed and unlimed sites (Teien et al. 2006), however, the decreases may not be directly related to the liming. The heterogeneity in effect shows that there are factors other than liming, not controlled for in the study designs, which have a significant impact on the abundance of fish. If other factors are limiting fish populations, for example other sources of pollution, are having a large impact on fish abundance and independently increase after calcium carbonate is applied the fish populations could decrease despite any potential benefit from decreased acidification. Equally, if a liming operation is ineffective at increasing pH it may have little impact on the fish abundance. Additionally, different fish taxa responded to liming in different ways. Where salmon and trout occurred together, salmon generally increased in abundance with liming but trout showed no clear change in abundance and decreased in several rivers. Overall, the impact of liming on the number of fish species is uncertain, with no clear evidence of an increase after liming.

In terms of invertebrates, the impact of liming depends on both the group of invertebrates and is highly variable between rivers. The limitations of the available studies mean that the impact of liming on invertebrate abundance is uncertain. Over all, the studies found no consistent evidence of an effect; however the few studies with least susceptibility to bias, BACI studies, suggest a decrease in abundance of invertebrates after liming. Unlike abundance, the average taxonomic richness of invertebrates across rivers increased with liming. However, there were still differences between studies such that among rivers the true effect ranges from positive to negative.

Unlike for all invertebrates, the abundance of acid sensitive invertebrates increased in all six studies in which they were investigated and the average impact was positive. However, the limited number of studies makes it hard to draw firm conclusions about the impact of liming on this group of organisms. The small number of studies and the variation in impact on abundance between studies meant the prediction interval still overlapped zero indicating that acid sensitive invertebrate abundance may decrease with liming in a minority of rivers. The number of acid sensitive invertebrate taxa also increased with liming, however, this result was driven by one study. There were only five studies investigating acid sensitive invertebrate taxonomic richness. The study which covered the largest number of streams actually reported a negative impact.

The impact of liming on ecosystems will not be restricted to fish and invertebrates but changes in the diversity and abundance of these organisms provide an indication of effects more widely in the ecosystem and at lower trophic levels. Additionally, liming may have sub-lethal impacts on biota and not cause direct death. The review does not cover studies on the physiological effects of liming, however, even sub lethal effects are likely to impact on the abundance and diversity of a population over time.

The perceived effectiveness of a liming project will depend on its objectives. If the aim is to increase fish abundance then the evidence from this report suggests that this is more likely to be achieved than if the intention is to increase invertebrate abundance. Hence, it is important to define the aim of a liming project at the outset. Within Europe, the Water Framework Directive requires the identification of what constitutes 'good ecological status' for each type of water body. These are known as reference conditions (EU 2000). Identifying what the ecosystem would have been like without degradation can be challenging but it is important that accurate reference conditions and classifications are used. In Sweden, past reference conditions have generally been set by classifying the water body using biological and chemical monitoring data, however, palaeolimnological studies have suggested several lakes that have been limed may actually have been naturally acidic (Norberg et al. 2008). In the UK, Water Framework Directive reference conditions have been identified by both monitoring and modelling (UKTAG 2007). The recommended tool for identifying acidification effects on aquatic biota in the UK is the River Invertebrate Classification Tool (RCIT; UKTAG 2007). In comparison to the RICT, abundance and species richness are relatively crude measures of the status of biological communities and ecosystems. An increase in abundance of invertebrates does not necessarily show restoration to 'natural' conditions; it may be due to an increase in acid tolerant species rather than species typical of non-acidified areas. The tools used in the Water Framework Directive are more sophisticated but they are relatively new, they vary across member states and so have not been used extensively in previous research. Nine of the rivers in the Norwegian liming survey used a specific acidification index to measure the impact of liming on invertebrates, however it was not used outside Norway. Hence, they could not be used as the measured outcomes in this review.

5.2. Reasons for variation in effectiveness

Predicting the likely effect of liming on a particular river is complicated by the large variation in effectiveness (heterogeneity) seen between studies and hence rivers. The differences suggest there are several environmental factors, not controlled for in the studies, which can alter how fish and invertebrates react to liming. Potential differences between liming operations that may influence effectiveness are also numerous and include the dose, type of liming and duration of liming. As the list is large, not all factors could be investigated with the limited number of robust studies found by this review. Out of the factors investigated (average duration of application, liming method, fish stocking, type of study and group of fish involved) the only factors that explained a significant amount of the variation were duration of liming and the group of fish involved.

The duration of liming is one factor that has been shown to predict differences in fish abundance. It can take time for populations to recover, especially if they have a long life cycle. It takes an average of between 4 and 6 years for Atlantic salmon (*Salmo salar*) to reach maturity depending on region (Hutchings and Jones 1998). Additionally, it can take time for species which become locally extinct due to acidification to recolonise. Therefore, maintaining liming programs in the long term is important. If the acidic deposition in an area is still outstripping the soil neutralising capacity as soon as a liming operation is stopped the pH will start to fall again (Clare and Hindar 2005), hence, in this situation the duration is most likely to be important compared to situations where the critical load is no longer exceeded. However, the observed effect of the duration of liming operation may not be simply due to the duration of liming, there are two other factors that need to be taken into consideration. As many of the studies were of BA design, decreases in acidic deposition over the time of the studies could have reduced acidification independent of liming. Decreases in acidic deposition are more likely to have produced a noticeable difference in longer term studies. Hence, the increase in abundance of fish with increased duration of liming may actually indicate the importance of decreasing atmospheric deposition rather than just the impact of liming. The scale of liming operations is also linked to the duration of the operation; small experimental studies are often of short duration whereas large scale national liming programs have been in continuous operation for much longer. Thus, the effect of average time of liming operation may be due to the scale of the liming project and the scale of liming project may be important for increasing the chances of success.

The response to liming is also affected by the organisms involved. As species differ in their physiologies and ecological requirements it is unsurprising that liming has varying impacts on them. Species do not live in isolation and interactions between species may result in indirect impacts on species abundance with liming. It has been suggested that trout numbers do not increase, or may even decrease, where salmon numbers increase, due to increased competition (Degerman and Appelberg 1992). Hence, the fish community composition at the outset of liming may also be important for determining the impact, along with the order of any recolonisation (Clair and Hindar 2005) and the presence of any stocking.

The duration of the liming operation and the species involved explains some but not all of the variations seen between studies. Many, possibly interlinked, factors are likely to be involved in the observed heterogeneity. The fact that the type of liming did not explain a significant amount of heterogeneity does not prove that the type of liming is not important or has no effect on the impact of liming. Instead, it may not have explained a significant amount of the heterogeneity due to only having a small sample and other factors causing more heterogeneity and variation in effect. Not all liming operations are likely to be equally effective and where the operations do not increase pH they are unlikely to allow recovery of acid sensitive species. Factors that could not be investigated in this review but may be important include the dose of calcium carbonate applied, the community of organisms already present, the chemical conditions in the stream including the presence of acid spikes, changes in land use or management and the precipitation of soluble aluminium salts when pH is increased by liming. Additionally, if acidification is not the main environmental pressure on a river then changes in other pressures, such as the impact of power plants, other pollution, aquaculture impacts or fishing pressures are likely to have a larger impact on the biota of the river than reductions in acidification by liming.

5.3 Review limitations

As with all systematic reviews it is important to understand the limits of the information it can provide. The conclusions are highly dependent on the studies found within the search. Over half of the studies included in the review did not have both control and baseline data; increasing the chances of bias within the original data. The potential of bias in the individual studies, overviewed in section 4.2 and individual study details in appendix B, decreases the confidence that can be placed on the overall mean effects calculated from the studies. Of particular note is the number of studies with only before and after data and no independent controls that may have been impacted by decreases in acid rain over the same period as the liming occurred. Isolating the impact of liming from decreasing acidification can be difficult but it is essential. If decreased deposition is causing recovery on its own, liming may not be necessary; liming is likely to have few long term benefits compared with natural recovery (Ormerod and Durance 2009).

The review is not only limited by the quality of included studies but also the variability between the studies and the absolute number of studies found. The review showed that there was high variability in the effects of liming amongst studies and understanding the reasons for these differences can help in making predictions about what will occur in other rivers. The differences observed are likely to be due to a large number of factors that vary between rivers but as there were only a limited number of studies, a restricted number of factors could be statistically investigated. Hence, the review can only provide limited information on what causes the observed variations in effect.

The other main limitation to the review is related to the question and aim of the review rather than the available studies. The review covered a limited question; “What is the impact of liming on fish and invertebrate diversity and abundance?” The limited question allowed the rigorous collection of all data on the topic and the statistical analysis of these data, however, not all factors relevant to the question of “should we lime?” were covered. In deciding on whether to implement a liming strategy other factors may need to be considered including the impact of liming on non-target habitats (including terrestrial habitats if catchment liming is implemented, Shore and Mackenzie 1993), whether factors other than acidification are impacting on the river (e.g. barriers to migration, Hesthagen et al. 2011), the cost of liming (Navrud 2001) and the political and social reasons for liming (Clair and Hindar 2005).

6. Reviewers' Conclusions

6.1. Implications for management / policy / conservation

Despite the limitations of the review there are several important policy implications from the findings:

- Liming is likely to increase overall fish stocks 1.7 times by comparison with the control locations (or before data) (confidence interval of the mean effect (response ratio) 1.3 - 2.1) but in any one river this is not guaranteed. Fish abundance is predicted to decrease with liming in around one-fifth of rivers.

- If liming targets salmon populations then there is more likely to be a positive impact. However, it is still possible to get a decrease in salmon numbers after liming. Liming is not the only factor affecting salmon populations as other factors also need to be optimal.

- Liming generally increases acid sensitive invertebrate abundance and number of taxa (mean effect (response ratio) = 1.96, CI=1.12-3.44) but the impact is variable and effectiveness not guaranteed, other factors modify the effect.

- The studies with least risk of bias suggest liming may decrease invertebrate abundance overall, however, variability in their findings and between different study designs means there is uncertainty in the impact of liming on invertebrates overall.

- There is insufficient high quality information on the impact of liming on invertebrates, especially studies with a BACI type design and replication. More research is required on this topic.

- Policy makers need to be clear about the aims of a liming project and what constitutes 'natural'/'desired' conditions. Available studies do not use ideal outcome measures for the context of the Water Framework Directive.

- Acid deposition is decreasing and recovery due to decreased deposition is hard to isolate from the impact of liming in BA studies (which were common); future studies need to include control sites.

6.2. Implications for research

The review findings have some important implications for future research. There are both implications for ways of improving future work and the areas in which future work is needed.

Implications for study designs

- Replicates should always be present within studies (either within or between rivers depending on the scale of the study, ideally both). In the studies found, replication was not always present preventing the calculation of uncertainty.
- The variances for each outcome should always be reported along with the scale over which the variance was calculated; whether the variance was within or between rivers needs to be stated. This was again lacking in some studies making calculations of the uncertainty in the findings difficult.
- There is a need for more studies to be of BACI design as they are generally at a lower risk of bias.

Implications for future research questions:

- The studies evaluated showed there is particular uncertainty about the impact of liming on invertebrate abundance; hence, more studies in this area are required.
- More research on confounding factors is also needed as there were large differences in the effectiveness of liming between studies but there were too few studies in this review to fully investigate the reasons.
- Abundance and species richness are crude measures of an ecosystem's quality; along with ongoing long term monitoring, there needs to be more research using outcome measures more appropriate to policy objectives, especially those linked to the Water Framework Directive.

7. Acknowledgements

Thanks to James Bussell and Diane Jones for their assistance in the early stages of this review. Also thanks to Barbara Livoreil for help with translating the French articles translated and to Tobias Vrede, Associate professor, Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences for compiling the Swedish dataset. This work was funded by the UK Natural Environment Research Council through Knowledge Exchange Grant NE/F009356/1.

8. Potential Conflicts of Interest and Sources of Support

None identified. The project is funded by the Natural Environment Research Committee UK (NERC).

9. References

Barlaup, B. T. (2004) *Vossolaksen – bestandsutvikling, trusselfaktorer og tiltak.* DN-utredning 2004-7. Direktoratet for naturforvaltning, Norway

Bostedt, G., Löfgren, S., Innala S. and Bishop K. (2010) Acidification remediation alternatives: exploring the temporal dimension with cost benefit analysis. *Ambio* 39 (1): 40-48

CEE (Collaboration for Environmental Evidence) (2010) *Guidelines for systematic reviews in environmental management. Version 4.0,* Collaboration for environmental evidence.

Clair, T. A. and Hindar, A. (2005) Liming for the mitigation of acid rain effects in freshwaters: a review of recent results. *Environmental Reviews* 13 (3): 91-128

Chestnut, L. G. and Mills, D. M. (2005) A fresh look at the benefits and costs of the US acid rain program. *Journal of Environmental Management* 77: 252–266

Degerman, E. and Appelberg, M. (1992) The response of stream-dwelling fish to liming. *Environmental Pollution* 78 (1-3): 149-55

Donnelly, A., Jennings E. and Allott N. (2003) A review of liming options for afforested catchments in Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy, Section B* 103B (2): 91-105

Downey, D. M., French, C. R. and Odom, M. (1994). Low cost limestone treatment of acid sensitive trout streams in the Appalachian Mountains of Virginia. *Water, Air, & Soil Pollution* 77(1-2): 49-77.

Edwards, P., Clarke M., DiGuiseppi, C., Pratap, S., Roberts, I. and Wentz R., 2002. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine* 21: 1635-1640

European Union (EU) (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy.

Evans, C.D., Cullen, J.M., Alewell, C., Kopácek, J., Marchetto, A., Moldan, F., Prechtel, A., Rogora, M., Veselý, J. and Wright, R. (2001) Recovery from acidification in European surface waters. *Hydrology and Earth System Sciences*, 5(3), 283-297

Evans, C. D., Reynolds, B., Hinton, C., Hughes S., Norris, D., Grant, S. and Williams, B. (2008) Effects of decreasing acid deposition and climate change on acid extremes in an upland stream. *Hydrology and Earth System Sciences* 12: 337–351

Fowler, D., Smith, R., Muller, J., Cape, J. N., Sutton M, Erisman, J. W. and Fagerli, H. (2007) Long term trends in Sulphur and Nitrogen deposition in Europe and the cause of non-linearities. *Water, Air and Soil Pollution: Focus* 7:41–47

Hesthagen, T. and Larsen, B. M. (2003) Recovery and re-establishment of Atlantic salmon, Salmo salar, in limed Norwegian rivers. *Fisheries Management and Ecology* 10 (2): 87-95

Hesthagen, T., Larsen, B. M., Fiske, P. (2011) Liming restores Atlantic salmon (*Salmo salar*) populations in acidified Norwegian rivers. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(2): 224-231

Hutchings, J.A. and Jones, M.E.B. (1998) Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Science* 55(Suppl. 1): 22–47

Kowalik, R.A., Cooper, D.M., Evans, C.D. and Ormerod, S.J. (2007) Acidic episodes retard the biological recovery of upland British streams from chronic acidification. *Global Change Biology* 13: 2439–2452

Kuylenstierna, J.C.I., Rodhe, H., Cinderby, S. and Hicks, K. (2001) Acidification in developing countries: ecosystem sensitivity and the critical load approach on a global scale. *Ambio*, 30 (1): 20-28

Liebich, T., McCormick, S. D., Kircheis, D., Johnson, K., Regal, R. and Hrabik, T. (2011) Water chemistry and its effects on the physiology and survival of Atlantic salmon *Salmo salar* smolts. *Journal of Fish Biology,* 79 (2): 502–519

Likens, G. E., Driscoll, C. T. and Buso, D. C. (1996) Long-term effects of acid rain: response and recovery of a forest ecosystem. *Science* 272: 244-246

Mant, R. Jones, D. Bussell, J. Jones,D. Godbold, D. Reynolds, B. Ormerod, S. and Pullin, A.S. 2010. Is 'liming' of streams and rivers an effective intervention for managing water quality to support fish and invertebrate populations? Collaboration for Environmental Evidence. www.environmentalevidence.org/SR76.html

Matejko, M., Dore, A.J., Hall, J., Dore, C.J., Błaś, M., Kryza, M., Smith, R. and Fowler, D. (2009) The influence of long term trends in pollutant emissions on deposition of sulphur and nitrogen and exceedance of critical loads in the United Kingdom. *Environmental Science and Policy* 12: 882-896

Menz, F.C. and Seip, H. M. (2004) Acid rain in Europe and the United States: an update. *Environmental Science & Policy* 7: 253–265

Moiseenko, T. I. (2005) Effects of acidification on aquatic ecosystems. *Russian Journal of Ecology*, 36(2): 93–102.

Navrud, S. (2001) Economic valuation of inland recreational fisheries: empirical studies and their policy use in Norway. *Fisheries Management and Ecology*, 8: 369-382

Nisbet, T.R. (2001) The role of forest management in controlling diffuse pollution in UK forestry. *Forest Ecology and Management,* 143 (1-3): 215-226

Norberg, M., Bigler, C. and Renberg, I. (2008) Monitoring compared with paleolimnology: implications for the definition of reference condition in limed lakes in Sweden. *Environmental Monitoring and Assessment* 146(1-3): 295-308

Ormerod, S. J., and Durance, I. (2009) Restoration and recovery from acidification in upland Welsh streams over 25 years. *Journal of Applied Ecology* 46:164-174.

Reynolds, B., Stevens, P. A., Brittain, S. A., Norris, D. A, Hughes, S. and Woods, C. (2004) Long-term changes in precipitation and stream water chemistry in small forest and moorland catchments at Beddgelert Forest, north Wales. *Hydrology and Earth System Sciences* 8: 436-448.

Sandøy, S. and Langåker, R. (2001) Atlantic salmon and acidification in southern Norway: a disaster in the 20th Century, but a hope for the future? *Water, Air, & Soil Pollution* 130(1-4): 1343-1348

Schindler, D. W., Mills, K. H., Malley, D. F., Findlay D. L., Shearer, J. A., Davies, I. J., Turner, M. A. Linsey, G. A. and Cruikshank, D. R. (1985) Long-term ecosystem stress: the effects of years of experimental acidification on a small lake. *Science* 288: 1395-1401

Singh, A. and Agrawal, M. (2008) Acid rain and its ecological consequences. *Journal of Environmental Biology* 29(1): 15-24

Shore, R.F. and Mackenzie, S. (1993) The effects of catchment liming on shrews *Sorex* spp.. *Biological Conservation* 64(2): 101-111

Teien, H-C, Kroglund, F., Salbu, B. and Rosseland, B. O. (2006) Gill reactivity of aluminium-species following liming. *Science of The Total Environment* 358 (1-3): 206-220

UKTAG (UK Technical Advisory Group on the Water Framework Directive) (2007) Recommendations on Surface Water Classification Schemes for the purposes of the Water Framework Directive. *available from http://www.wfduk.org/*

Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor Package. *Journal of Statistical Software* 36(3): 1-48

Watt, W. D. (1987) A summary of the impact of acid rain on Atlantic salmon (*Salmo salar*) in Canada**.** *Water, Air, and Soil Pollution* 35(1-2): 27-35

Watt, W. D., Farmer, G. J. and White, W. J. (1983). Studies on the use of limestone to restore Atlantic salmon habitat in acidified rivers. *Lake and Reservoir Management* 1(1): 374-379.