Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Results	Notes
Authors: Anon (ADAS report)	**Source population:** Semi-natural grasslands, particularly hay meadows	**Methods of allocation:** NA	**Primary outcome measures:** NA	review was unimproved grassland of nature conservation value, particularly hay meadows there was relatively little information available for this grassland type, so it was necessary to consider data from more intensively managed grassland. They reported that it is likely that lower nutrient inputs will increase species diversity, and that there is no evidence to indicate differences in the effectiveness of inorganic compared to organic fertilisers of the same NPK analysis in their effect both on grass growth and on sward botanical	**Limitations identified by author:** Authors state that the effects of organic manures on conservation objectives has not been studied
Year: 1993(?)	**Eligible Population:** NA	**Intervention description:** NA	**Secondary outcome measures:** NA		**Limitations identified by review team:** Barely any mention of species diversity in the report
Aim of study: To summarise information in the nutrient requirements of grassland, the main sources of nutrients and commercially available fertilisers and differences in their behaviour	**Inclusion & exclusion criteria:** Not given	**Control / comparison description:** NA	**Follow-up periods:** NA		**Evidence gaps and/pr recommendations for further research:**
Study design: 4 - a review of existing literature and expert opinion (the latter only for any mention of plant species diversity)	**Setting:** UK	**Sample sizes:** NA	**Methods of analysis:** NA		**Sources of funding:** Not reported
Quality Score: -		**Baseline comparisons:** NA			
External validity: -		**Study sufficiently powered:** NA			

Study question	a - effect of fertiliser on species diversity				
Citation	Comparative Effects of Organic-Based and Inorganic Fertilisers, and Organic Manures for Unimproved Semi Natural Grasslands. ADAS Report 1993				
Study category	4 - Review/expert				
Assessed by	Kate Fagan 26/11/12				
Section 1: Theoretical approach		Comments:			
1.1 Is a qualitative approach appropriate?	Not sure				
Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings?					
Could a quantitative approach better have addressed the research question?					
1.2 Is the study clear in what it seeks to do?	Clear				
For example:					

is the purpose of the study discussed – aims/objectives/research questions?					
is there adequate / appropriate reference to literature?					
are underpinning values / assumptions discussed?					
1.3 How defensible / rigorous is the research design / methodology?	□ Indefensible	No methodology or study design - literature review			
For example:					
is the design appropriate to the research question?					
is a rationale given for using a qualitative approach?					
are there clear accounts of the rationale for sampling, data collection and data analysis techniques used?					
Is the selection of cases / sampling strategy theoretically justified?					
Section 2: Study Design					
2.1 How defensible / rigorous is the research design / methodology?	□ Indefensible	No methodology or study design - literature review			

For example:					
is the design appropriate to the research question?					
is a rationale given for using a qualitative approach?					
are there clear accounts of the rationale for sampling, data collection and data analysis techniques used?					
Is the selection of cases / sampling strategy theoretically justified?					
Section 3: Data Collection					
3.1 How well was the data collection carried out?	□ Inappropriately				
For example:					
are data collection methods clearly described?					
Were the appropriate data collected to address the research question?					
Was the data collection and record keeping systematic?					
Section 4:Trustworthiness					

4.1 Is the role of researcher clearly described?	□Not Applicable				
For example:					
has the relationship between the researchers and intervention group been adequately considered?					
4.2 Is the context clearly described?	□Not Applicable				
For example					
Were observations made in a sufficient variaty of circumstances?					
Was context bias considered?					
4.3 Were the methods reliable?	□ Not Sure				
For example:					
Was data collected by more than one method?					
is there justification for triangulation or for not triangulating?					
do the methods investigate what they claim to?					
Section 5: Analyses					
5.1 Is the data analysis sufficiently rigorous?	Not applicable				
For example:					

is the procedure explicit?					
how systematic is the analysis, is the procedure reliable?					
is it clear how the themes and concepts Were derived from the data?					
5.2 Is the data 'rich'?	Not applicable				
For example:					
how well are the contexts of the data described?					
has the diversity of perspective and content been explored?					
are responses compared and contrasted?					
5.3 Is the analysis reliable?	Not applicable				
For example:					
did more than one researcher theme and code data?					
if so how Were differences resolved?					
Were negative / discrepant Results addressed?					
5.4 Are findings convincing?	☐ Convincing				
For example:					
findings clearly presented?					

finding internally coherent?					
Extracts from original data included?					
data appropriately referenced?					
reporting clear and coherent?					
5.5 Are the findings relevant to the aims of the study?	□ Partially relevant				
5.6 Conclusions					
For example:	Not applicable				
how clear are the links between data interpretation and conclusions?					
are the conclusions plausible and coherent?					
have alternative explanations been explored and discounted?					
does this enhance understanding of the research topic?					
are the implications of the research clearly defined?					
is there adequate discussion of the limitations encountered?					
Section 6: Ethics					

6.1 How clear and coherent is the reporting of ethics?	Not applicable				
For example:					
have ethical issues been taken into consideration?					
are they adequately considered?					
have the consequences of the research been considered?					
Was the study approved by an ethics committee?					
Section 7: Overall Assessment					
As far as can be ascertained from the paper, how well was the study conducted?	□ -				
For example:					
are data collection methods clearly described?					
Were the appropriate data collected to address the research question?					
Was the data collection and record keeping systematic?					

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	ADAS
	Year	1996
	Aim of study	To monitor populations of breeding waders and yellow wagtails in land under agreement in the Pennine Dales ESA and detect population changes between 1991-1995. The main focus of the yellow wagtail work was to examine the timing of nesting and fledging in relation to the timing of grass cutting.
	Study design	2 Monitoring study
	Quality score	+
	External validity	+
Population and setting	Source population	Yellow wagtail population in Pennine Dales ESA
	Eligible population	Yellow wagtail populations within hay meadows in Northern England

	Inclusion and exclusion criteria	
	Setting	Teesdale, Weardale in 1991. In 1992 the study was extended to Upper Wensleydale, Garsdale, Mallerstang, Upper Eden valley, Grisedale, Rawthey Valley, Dentdale and Deepdale
Methods of allocation to intervention/control	Methods of allocation	Teesdale and Weardale were selected on the basis that theu were known to be dales with large numbers of yellow wagtails – ensuring maximum possible opportunity to monitor nests and determine fledging dates. The study was extended in 1992 to the Western dales listed above which in contrast to Teesdale and Weardale were thought to have a small number of yellow wagtails. As much as possible of the dales were studies and the location and number of pairs were established .
	Intervention description	
	Control/comparison description	Meadow nest sites were identified in both agreement and non-agreement land except in 1992 in Teesdale and Weardale. In the Western dales where the bird populations are small, both pasture and meadow sites were monitored.
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	1991 – Teesdale and Weardale 1993 – Western Dales
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Field visits started during the second week in May and continued until all breeding activity had stopped usually by August. In the phase of the survey all dales were visited on at least a weekly basis. Once pairs had been established they were observed on at least a weekly basis, following a fixed route. At each visit, each potential nest pair was allocated on the basis of behaviour to one of five stages. Nest sites were recorded as vacated after two successive visits noted no further activity.
	Secondary outcome	

	measures	
	Follow-up periods	Sites were monitored in 1991, 1992 , 1993 and 1995 in Teesdale and Weardale Sites were monitored from 1993 to 1995 in the Western Dales
	Methods of analysis	Observational data were simply presented - no statistical analysis was undertaken.
Results		The main focus of the yellow wagtail work was to examine the timing of nesting and fledging in relation to the timing of grass cutting. They found a clear preference for yellow wagtails to nest to meadows within the Pennine Dales ESA. They also found that peak fledging date in Dales is the last week of June, with approximately 70% of birds fledging prior to the 7th July. Over the survey period a quarter of nests failed due to cutting, with the impact of cutting in any one year varying with both the timing of the breeding season and the timing of cutting, this vary widely with spring temperature and rainfall. The 8th July cutting date for the ESA falling just after the peak fledging period is judged to offer considerable protection for the breeding population on agreement land.
Notes	Limitations identified by author	Response of the overall population of breeding yellow wagtails to the more favourable meadow cutting regime is not yet clear from these results.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Defra

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	ADAS (1996) Bird monitoring in the Pennine Dales 1991 – 1995. ADAS unpublished report
Study Design Category	2
Assessed by & when	CE Pinches, 12th December 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Yes the Yellow Wagtail, its preferred habitat requirements are well described.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Yes, survey coverage is over a large geographic area within the Pennine Dales ESA
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Selection is targeted not random but contrasts high and low density areas of yellow wagtail populations and as such is representative of the population overall.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NR	Comments: Not relevant
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes – study sought to look at the impact of management practices on bird breeding success and ultimately populations, in particular to look at impact of meadow cutting on fledging.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NR	Comments: Not relevant
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐NR	Comments: Factors other than cutting date not considered.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Yes objective intensive observational measures
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐+	Comments: Yes with the exception of non-agreement nest sites in 1992 in Teesdale and Weardale.
3.3 Were all important outcomes assessed? Were all important positive and negative	☐++	Comments: Yes

effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: Not applicable
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: No not to assess long term effects of later hay cuts on yellow wagtail populations as a whole

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NA	Comments: Presentation of observational survey data
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: No, Not Applicable description of monitoring results only.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: Difference between agreement and no-agreement could have been tested where number of nesting pairs were sufficient.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NR	Comments: No
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐+	Comments: Yes

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Aerts, de Caluwe & Beltman
	Year	2003
	Aim of study	To test the hypothesis that increased supply of a growth limiting nutrient (either N or P) promotes biodiversity in grasslands
	Study design	2
	Quality score	+
	External validity	-
Population and setting	Source population	Riverine grassland site over sandy clay in Netherlands
	Eligible population	Grassland belongs to alliance Arrhenateretalia, under which northern hay meadows also sit.
	Inclusion and exclusion criteria	Study also looked at a peat grassland site results for which are ignored for purpose of this review
	Setting	Netherlands
Methods of allocation to	Methods of allocation	Three permanent plots 4 x 4 m. Each plot was divided into 4 subplots of 2 x 2m.

intervention/control	Intervention description	Unfertilised (O), Fertilised with N (N, fertilised with P (P) and fertilised with both and N and P (N+P). N and P applied in granular form twice a year in form of Nh4No3 (10g N m2/yr) and NaH2PO4 (5g P/m2/yr).
	Control/comparison description	Unfertilised control.
	Sample sizes	Replication 3 x
	Baseline comparisons	Initial vegetation survey undertaken in 1985 Braun Blanquet scale - species cover percentages were subsequently estimated from this scale
	Study sufficiently powered	Variability after 11 years was found to be high for N mineralisation and Extractable P measures. and authors report that this combined with relatively low replications means that there may have been relatively low power to detect sig diff for this element o fthe experiment. Power for other elements deemed to be satisfactory.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Soil sampling - 2x 10cm depth cores randomly sampled within each sub plot – organic matter content (LOI), Total C, N and P content, N:P mass ratios Net N mineralisation and P release - soil sampled in 1996, two weeks after first annual fertilisation one sample taken back to labs immediately, one sample left to incubate in situ and analysed after 6 weeks. Vegetation – Species % cover estimates made in 1996 Peak biomass – end of July.
	Secondary outcome measures	
	Follow-up periods	Outcome measure recorded following 11 years of nutrient additions.
	Methods of analysis	GLM - with relevant transformation on non normally distributed datasets. Effects of treatments tested using Tukey's HSD test.

		Soil and vegetation nutrient data analysed using regression. Multiple linear regressions were used to relate vegetation characteristics to the soil nutrient parameters.
Results		**Soil nutrients** – The N:P ratio of 6 apparent in the control is a strong indicator of N limited plant growth, in contrast P was not limiting in this grassland. After 11 years there were no significant difference between fertiliser treatments on total nutrient pools in the soils. Nitrogen mineralisation was significantly increased in treatments were N was added. Similarly P release was significantly increased in treatments were P was added ($p < 0.05$ in both cases). If these results are extrapolated over an entire growing season the findings indicate that N and P fertilisation can therefore have a significant impact on the annual flux and cycling of an apparently small pool of labile N and P compounds. **Vegetation** There was a significant increase in peak standing biomass when N and P were supplied. N removal by hay making compensated for atmospheric N input (about 4g Nm2/yr) but did not compensate for the added nutrients in the fertilise treatments. P removal by hay making in the P containing treatments was only 20% of that applied. The relatively higher N removal is due to the higher allocation of N to above ground parts of plans as N is the main nutrient of photosynthetic tissues. Species richness was highest under non fertilised control (22 species per plot). Addition of N resulted in a reduction in species diversity and evenness with a strong reduction in the number of legumes and a strong increase in grasses. After 11 years of treatments, species diversity and evenness were strongly determined by N mineralisation and to a lesser extent by total soil N and extractable P respectively. Addition of the growth limiting nutrient for this grassland(i,e N) lead to lower species diversity and higher biomass. No such effects were observed with additions of the non-limiting nutrient (P). Even at relatively low biomass, addition of N may lead to a reduction of species diversity due to extinction of legumes and other low statured species which may be outshaded due to more intense competition for light from tall grasses. At balanced N:P ratios competitive

		interactions may be more important determinants of biodiversity than differentiation of nutrient acquisition strategy. Another possible explanation is that once species had been lost there may be insufficient propagules of species to re-colonised the nutrient balanced treatments.
Notes	Limitations identified by author	Possible insufficient power on N mineralisation and P extractable measures.
	Limitations identified by review team	Study focuses on vegetation at t = 0 and t = 11 i.e only vegetation change at end point of treatments, therefore it is impossible to determine trends over time in species change/colonisation and extinction which may give a better indication of whether propagules are a limiting factor or not which compounds the loss of diversity resulting from competition for light under increased N.
	Evidence gaps and/pr recommendations for further research	Possible role of biotic (propagule) constraint as an explanation should be explored alongside continued nutrient additions..
	Sources of funding	Unclear

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	a)What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?
Study Citation	Aerts, R., de Caluwe, H., & Beltman, B. (2003). Is the relation between nutrient supply and biodiversity co-determined by the type of nutrient limitation? *Oikos, 101*, 489-498.
Study Design Category	2
Assessed by & when	C.E. Pinches, 6th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Plant community and soil chemical properties and historic management all well described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Manner by which experimental sites were selected is not described but is not random
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	Comments: No information is provided on how plot location was determined (e.g. randomly) within each meadow sampled.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐+	Comments: It is not clear whether treatments were allocated randomly to the subplots within the three replicate plots.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: All 16 treatments well described and would enable replication. The comparison is appropriate and is an untreated control.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes, broadly although rates of application for both N and P were quite low.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Yes
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes, a traditional hay cutting and grazing regime were applied across all treatments. The details of this are well described.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐-	Comments: Similar riverine grasslands found in England but less strictly relevant to upland hay meadow context.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Rates of application are lower than the norm but are equivalent to rates previously allowed under ESA tier 1.

Section 3: Outcomes		
3.1 Were outcome variables/measures		Comments: Both - Subjective botanical assessments -

reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	% cover of each species present. Objective - soil sampling for soil nutrients and measures of microbial community structure.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes botanical and soil measures are appropriate.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes, 11 years, the experiment would have benefitted from annual botanical monitoring to determined change over time in botanical composition, colonisation and extinction events within plots.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Yes, although not clear if block has been used as a variable.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐+	Comments: In most elements but data variability in N mineralisation and Extractable P indicate experiment

A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		may be insufficiently powered to detect sig effects in these outcome measures.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes.
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Yes.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes, but only as P<0.05 no actul P values provide.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Yes
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐-	Comments: The findings are valid but are less directly relevant to MG3 meadows.

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Askew, D.R. (1994). Pennine Dales ESA: grassland management and and nature conservation interest. In: R.J. Haggar, & S. Peel (Eds.), *Grassland management and nature conservation: Proceedings of a joint meeting between the British Grassland Society and the British Ecological Society held at Leeds University 27-29 September 1993.* (pp. 179-184).
Study Design Category	2
Assessed by & when	CE Pinches 25th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Yes, comprehensive description of meadows and agricultural context in which they exist
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐ ++	Yes
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Fields were randomly selected from those under ESA agreement , except in the case of 25 fields whose selection was dependent upon the co-operation of farmers who were not in the scheme.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments: NA Correlative study
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Yes data for 20 management variables were collected
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Acknowledged but not controlled for, for example fertiliser use and cutting date.
2.5 Is the setting applicable to the UK?	☐+	yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Two outcome measures were used in the analysis Namely 1:field mean % of species quadrat classified as stress tolerant, following Grime's functional model – this was used as a proxy for nature conservation value (fields with a high percentage cover of these stress tolerators having high conservation value). Target thresholds of 15 % and 8% cover were used to a reasonable split between fields of high and low interest. Presence of wood crane's bill was also used as an outcome measure to indicate high value. The study may have benefitted from using a wider range of outcome measures
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions	☐+	Yes

likely to have been identified?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	It may have been useful to look at wider suite of botanical outcomes – e.g. fit to NVC type.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Difficult to test whether the effectiveness of using % stress tolerator species cover as proxy.
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Study is in effect a baseline looking at associations between management practices and nature conservation interest.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐NA	

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NA	NA but no of meadows included in study is large 307.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Yes 20 management variables were used. .
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	A non parametric Chi-Squared Automatic Intercation Detector segmentation technique was used – the end result of the analysis is a hierarchy of significant (P) management variables identifying the management associated with high or low conservation interest,
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐-	No

5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Yes, as a general description of management factors associated with high or low conservation interest.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Yes

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Askew
	Year	1993
	Aim of study	To analyse botanical and management data from a sample of meadows in the ESA in 1987 and 1989 to identify management associated with meadows of high or low conservation interest.
	Study design	2 (Correlative study)
	Quality score	+
	External validity	+
Population and setting	Source population	Meadows in Pennine Dales ESA
	Eligible population	As above
	Inclusion and exclusion criteria	-

	Setting	
Methods of allocation to intervention/control	Methods of allocation	NA Correlative/associative study
	Intervention description	NA
	Control/comparison description	NA
	Sample sizes	307 meadows
	Baseline comparisons	No – this represents the baseline
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Two outcome measures were used in the analysis Namely 1:field mean % of species quadrat classified as stress tolerant, following Grime's functional model – this was used as a proxy for nature conservation value (fields with a high percentage cover of these stress tolerators having high conservation value). Target thresholds of 15 % and 8% cover were used to a reasonable split between fields of high and low interest. Presence of wood crane's bill was also used as an outcome measure to indicate high value
	Secondary outcome measures	None used.
	Follow-up periods	NA
	Methods of analysis	Chi squared analysis of categorical data applied provides hierarchy of significant management variables identifying management associated with high or low conservation interest.

Results		Factors important in distinguishing meadows with conservation interest were: • Hay making as opposed to silage • Absence of improved drainage • Lower fertiliser input • Later cutting date • Also some lime applications (not no lime). More stress tolerating species associated with pastures not meadows (but not higher species richness) For Wood crane's bill, meadow grazing regime is identified as most factor most significantly associated with its presence (long winter and moderate spring grazing). For Yellow rattle, later cutting is identified as a significant factor but after herbicide and lime application.
Notes	Limitations identified by author	Potentially confounding effects of some of the management variables – e.g nutrient inputs and hay cutting date. The author notes that the data do not cover important environmental variables such as soil type.
	Limitations identified by review team	Whilst the approach used makes sense it is difficult to know how effective a proxy for nature conservation value % cover of stress tolerators is and how justified the targets set are in categorising low and high value sites.
	Evidence gaps and/pr recommendations for	MAFF

	further research	
	Sources of funding	

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a)What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ? c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Baines, D.
	Year	1990
	Aim of study	1. To assess the relative role of predation, food limitation and clutch destruction due to agricultural activity in determining breeding success of lapwings on upland grassland
	Study design	2
	Quality score	++
	External validity	++
Population and setting	Source population	Lapwing populations, Eden valley Cumbria and Teesdale Co.Durham
	Eligible population	Lapwing populations in Northern England
	Inclusion and exclusion	Fields were classified as either agriculturally improved or unimproved.

	criteria	
	Setting	Teesdale, between Alston and Langdon Beck , Co Durham and Head of Eden Valley near Brough and Kirkby Stephen, Cumbria
Methods of allocation to intervention/control	Methods of allocation	NA Observational study.
	Intervention description	Observational study – sample split between improved and unimproved usually paired sites.
	Control/comparison description	States that where possible unimproved meadows were paired with adjacent improved ones.
	Sample sizes	18 unimproved meadows and 15 improved meadows
	Baseline comparisons	NA
	Study sufficiently powered	Yes
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Lapwing nests - clutch size recorded - nests visited every 4 days Determination of hatching and no of chicks, % of eggs laid that hatched including replacements. Determination of clutch loss – either due to agricultural activity or predation. No of surviving chicks - broods examined every 4 days Density of lapwings 4 counts between April and end of May) Growth of chicks – chicks weighed every 4 days between 5 and 30 days.
	Secondary outcome measures	Sampling of main invertebrate constituents of lapwing diet (April), e.g earthworm, tipulid larvae (late April/May), surface active invertebrates (mid March to end of October)
	Follow-up periods	1985 - 1987

	Methods of analysis	Chi square and t tests.
Results		Baines (1990) reported that 22% of lapwing clutches laid on improved meadows were destroyed by farm machinery compared to 8% on unimproved meadows ($p<0.02$). Reduced lapwing productivity on improved meadows is attributable to more intensive management resulting in higher clutch loss to agricultural activities and the production of a faster growing sward that leaves insufficient time for replacement clutches.
Notes	Limitations identified by author	-
	Limitations identified by review team	Potentially considerable variability in within field operations not accounted for by crude improved/unimproved split
	Evidence gaps and/pr recommendations for further research	-
	Sources of funding	NERC research studenstship

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ? c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Baines, D 1990. The roles, of predation, food and agricultural practice in determining the breeding success of the lapwing (*Vanellus vanellus*) on upland grasslands. *Journal of Animal Ecology* 59: 915-929.
Study Design Category	2
Assessed by & when	CE Pinches, 10th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: Yes agriculturally improved or unimproved grasslands
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Yes, significant sample
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Selection was subjective.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	NA	Comments: Observational study NA
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Experiment sought to simply understand the relative role of predation, food limitation and clutch destruction due to agricultural activity in determining breeding success of lapwings on upland grassland
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐+	Comments: Distinction between improved and unimproved grassland made on basis of various management interventions, not on botanical composition.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Comments: There are potentially confounding factors, for example in field operations which could take place across categories in the sample.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Objective, reliable with accepted estimates.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative	☐+	Comments: Yes

effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: Yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: Yes, 3 years
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: Yes

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐ +	Comments: Sample size ok.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: No
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: Broadly ok.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?		Comments: Baines (1990) reported that 22% of lapwing clutches laid on improved

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	meadows were destroyed by farm machinery compared to 8% on unimproved meadows (p<0.02). Reduced lapwing productivity on improved meadows is attributable to more intensive management resulting in higher clutch loss to agricultural activities and the production of a faster growing sward that leaves insufficient time for replacement clutches.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++	Comments: Yes.

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: __Upland______________________

Name of Review Sub-topic (if any): ___Upland hay Meadow______

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Beintema, A. J., & Müskens, G. J. D. M. (1987) Nesting success of birds breeding in Dutch agricultural grasslands. Journal of Applied Ecology, 24, 743-758
Study Design Category	2
Assessed by & when	D Martin 22/11/12

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: The extent of Dutch meadow grasslands with breeding bird populations. Distinguished by high water table but increasingly subject to drainage and intensification. Not described in terms of vegetation.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: Source and eligible population assumed to be largely the same.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Methods of identification of study fields not described, as originally selected for a number of research projects. Fields with breeding birds chosen. It is likely that they were selected to be representative, and form a large sample (18 000 nest records). Possible bias in that nests lost before they could have been found in the survey are not included. Adjusted for by method of calculating survival rate.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Survey approach rather than comparison study. All fields selected for presence of breeding birds. Large sample size.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Explanatory variables are predation rates based on regular nests visits and observations of damage, and trampling rates influenced by cattle density and field size. The latter was explored in previous studies, but eventually reduced to survival per animal per ha per day. Methods of calculating survival rates adapted from Mayfield (1961, 1975)
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: Concerns of increased predation through nest marking – surveyors thought this not to be the case. Where it was thought predators had learned to follow surveyors trail, these nests were excluded.
2.5 Is the setting applicable to the UK?	☐+	Comments: Similar range of bird species. The setting is more applicable to lowland grassland and grazing marsh rather than upland hay meadow (smaller fields, walled, lower water table).

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measures subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Based on field observations of nest damage. May be some difficulty in distinguishing hatching from predation, and nests in the laying phase from partly predated abandoned clutches. Calculated predation rates early in the season are low, so latter is less of an issue.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions	☐++	Comments:

likely to have been identified?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Daily survival rates for 17 species and seasonal variation in predation rates for four most numerous. Trampling losses for different species under different grazing regimes.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: Survey approach rather than a trial. Nests were checked once or twice per week until
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Yes, all effects are measurable within nesting season. Data from a number of different seasons used.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments: No power analysis presented. Study based on a large number of observations.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: As well as those mentioned in 2.2 other variables used include species classification of nest hiding (subjective, by surveyors), and four types of livestock/ grazing system. Nest survival rates also tested against two classes of cattle density and three field size classes.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Spearman's rank correlation of nest hiding and predation, and survival against cattle density and field size. Models of interspecific swamping effects on predation rates
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for	☐++	Comments: p values given for correlations and regression models, and 95% confidence intervals for graphical comparisons.

the effect estimates given or calculable?		
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	Comments: Correlative census type study. Large sample but some lack of clarity over selection. A number of possible factors affecting predation rate and trampling investigated, within the limitations of the study approach.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Would be highly generalisable to Dutch grasslands, and possibly lowland grazing marsh in UK. Relevance to UHM reduced due to differences in farming systems, although effect of field size was shown not to be significant. Although small field classes were considered (<2.5 ha) other landscape factors might have an effect.

Mayfield, H. F. (1961) Nesting success calculated from exposure. Wilson Bulletin 36, 255-261.

Mayfield, H. F. (1975) Suggestions for calculating nest success. Wilson Bulletin, 73, 456-466.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Upland Hay Meadow
Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Beintema, A. J., & Müskens, G. J. D. M.
	Year	1987
	Aim of study	To analyse nest loss, and identify its significance in the population dynamics of meadow-bird species (wading birds).
	Study design	2
	Quality score	++
	External validity	+
Population and setting	Source population	The extent of Dutch meadow grasslands with breeding bird populations. Distinguished by high water table but increasingly subject to drainage and intensification. Not described in terms of vegetation.
	Eligible population	Source and eligible population assumed to be largely the same.
	Inclusion and exclusion criteria	Comments: Methods of identification of study fields not described, as originally selected for a number of research projects. Fields with breeding birds chosen. It is likely that they were selected to be representative, and form a large sample (18 000 nest records).

	Setting	Grasslands, generally part of dairy farming systems, in the Netherlands.
Methods of allocation to intervention/control	Methods of allocation	Survey approach
	Intervention description	Observational study of nesting success – effects explored are predation and trampling by livestock.
	Control/comparison description	Survey rather than controlled trial.
	Sample sizes	No indication of number or area of fields surveyed, but around 18 000 nests observed over 10 years.
	Baseline comparisons	NA
	Study sufficiently powered	No power analysis presented. Study based on a large number of observations.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Daily survival rates for 17 species and seasonal variation in predation rates for four most numerous.
	Secondary outcome measures	Effects of inter-specific swamping (effect of presence of different bird species on predation rates). Trampling losses for different species under different grazing regimes.
	Follow-up periods	Yes, all effects are measurable within nesting season. Data from a number of different seasons used.
	Methods of analysis	Daily survival rates calculated. Spearman's rank correlation of nest hiding and predation, and survival against cattle density and field size. Models of inter-specific swamping effects on predation rates. Re-nesting was modelled using a previously developed model.
Results		Survival rates during laying were lower than in the incubation phase. Overall, predation

		rates were high (half of all nests), but nesting success was high due to re-nesting. Predation therefore not considered a major threat. Losses to predation were higher than to trampling for lapwing, godwit and oystercatcher, but for redshank over 50% of nest losses were due to trampling. The impact of management exceeds losses due to predation, particularly at high stock densities. Young cattle were the worst tramplers for most bird species, especially when considered in terms of grazing equivalents (LU). Sheep did little harm per individual, but damage increases with stocking density. However the reduction in nesting success with increased density is less than for the equivalent cattle grazing pressure. There was little evidence of a significant interaction between grazing and predation for four species investigated. The probability of surviving mowing is zero. Lapwing may abandon a nest when vegetation becomes too tall, which could affect replacement clutches, which were shown to be an important part of the productivity of meadow birds. Other studies have shown that nesting season may end earlier in dry conditions, which can be exacerbated with improved drainage, and predation may also be facilitated by dry conditions. **(Relevance to UHM- Sheep spring grazing poses a moderate trampling risk, higher for redshank than lapwing. Where spring grazing is present, overall success likely to be higher in wet meadows, or those of low productivity or later closing, where re-nesting is most likely. Curlew were not considered in this study)**
Notes	Limitations identified by author	Limitations of re-nesting model, particularly the effects of management, or state of drainage.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for	More research into the probability of re-nesting especially later in the season.

	further research	
	Sources of funding	

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?

Study details	Authors	Breeuwer et al.
	Year	2009
	Aim of study	To assess the effectiveness of Dutch agri-environemht scheme in maintaining (and increasing) breeding bird species of meadows by analysing changes in the density of these species on land inside and outside agreement over a 12 year period.
	Study design	2 Correlative/Observational
	Quality score	2+
	External validity	+
Population and setting	Source population	Breeding birds of meadow grassland
	Eligible population	
	Inclusion and exclusion criteria	Grasslands included only in they were included within areas recognised by Dutch government as being sufficiently favourable meadow birds Selected within this area pairs of sites with and without management agreements, that had: (1) equal areas, (2) were

		located within 1 km of each other and more than 1 km from other selected sites, (3) had the same soil type and groundwater level and (4) were located in landscapes with a similar structure and at similar distances from roads, buildings and tree lines. In addition only included those pairs where bird counts had been performed in at least two years preceding the start of the agreement and two years after the start of the agreement (including the year in which the agreement started) and where these counts at the sites with and without contract had been performed in the same year.
	Setting	Twelve pairs of sites were located in the core meadow bird regions of the Netherlands; the others in smaller areas of suitable habitat elsewhere.
Methods of allocation to intervention/control	Methods of allocation	As above (see inclusion/exclusion criteria)
	Intervention description	Main intervention being investigated is postponement of mowing and other disturbing agricultural activities, such as manure application, to the end of May or June to reduce chick and egg mortality.
	Control/comparison description	Grasslands within or outside agri-environment agreements
	Sample sizes	28 pairs of sites for oystercatcher and black-tailed godwit. 26 and 24 pairs of sites, respectively f or lapwing and redshank
	Baseline comparisons	Yes, at least 2 year baseline for all samples.
	Study sufficiently powered	Yes.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Bird territories were surveyed during five field visits between 15 March and 15 June. The location of territories was assessed on the basis of the observations of nests, chicks and adult birds following the guidelines for the Breeding Bird Monitoring project in the Netherlands (van Dijk, 1996) which resembles the method used by the Common Bird Census in the UK.

	Secondary outcome measures	
	Follow-up periods	12 years
	Methods of analysis	GLM analysis were used to compare : – territory densities of the four bird species before and after the start of the agreement in the areas with and without contract, – the change in densities over time in control and managed areas before and after the start of the contracts, thus looking at effects on population development rather than density, added years before and after the start of AES agreement as a covariate. – interaction between the effects of management agreement and within-pair distance between – control and management sites, to control for the possible overflow of birds to neighbouring areas.
Results		Oystercatcher densities didn't differ between areas with and without management agreement, either before or after the start of the agreements The agreements did not have positive effects on the number of black-tailed godwits, and even had significant negative effects on the number of lapwings and redshanks relative to their numbers on control fields. 1. Improved conditions for reproduction do not result in increased local densities, but in an increased overflow of birds to neighboring areas. However , including the distance between managed and control fields in the statistical model did, however, not change the main results of the analysis (i.e. no or even negative effects of management agreements). In addition, the significant decline of the lapwing numbers on the managed fields relative to the control fields contradicts this hypothesis and suggest other factors may be causing the decline 2. Prescribed and paid management measures are not sufficient. In addition to the prescribed postponement of the mowing date it is probably necessary to raise

		groundwater levels and to reduce fertilization to allow for the development of an open vegetation structure that might increase chick survival to sufficiently high levels. i.e Other aspects of meadow management, namely drainage and nutrient inputs are likely to be indirectly affecting bird densities by reducing both the total amount of invertebrate prey available to the birds and it's accessibility.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	There is a lack of evidence as to where the young birds recruit into the breeding population. Need to determine if in the apidly changing agricultural landscape the environmental cues that birds use for the selection of breeding habitats are still those that are most appropriate
	Sources of funding	Birdlife The Netherlands and the Office for Environmental Outlooks

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?
Study Citation	Breeuwer, A., Berendse, F., Willems, F., Foppen, R., Teunissen, W., Schekkerman, H., & Goedhart, P. (2009). Do meadow birds profit from agri-environment schemes in Dutch agricultural landscapes? *Biological Conservation, 142*, 2949-2953.
Study Design Category	2
Assessed by & when	CE Pinches, 15th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Yes briefly .
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes, care in selecting sites was taken to ensure that bird populations to be sampled representative of populations on grassland in AES and outside it.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Well described and exclusion and inclusion criteria were explicit.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Exposure to postponed mowing date based on selecting paired meadows either in or outside AES agreement.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes, put simply it tests whether postponement of mowing date alone is sufficient to increase density.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐ NR or ☐-	Comments: Presumably there were differences in other management interventions within and across the within and outside AES categories – none were reported.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐ NR or ☐-	Comments: Potentially confounding factors not reported.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Yes, standard and accepted methodologies applied.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative	☐+	Comments: The inclusion of proxy measures of prey availability/accessibility i.e penetrability of soil, measure of grassland productivity/sward density

effects assessed?		would have helped determine other key management factors affecting bird density.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	Comments: No not always but this was controlled for in analyses.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐ +	Comments: Yes, 12 year study.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐ +	Comments: Yes, sample size and power fine.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: Yes, but none in relation to other management factors which may be significant.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: Yes
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally		Comments: Yes, show that Dutch AES have not been

valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	successful in maintaining breeding bird densities for meadow species but can only speculate on additional factors influencing this.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Broyer, J. **Year:**	**Source population:** No description of grasslands given other than low/medium/high altitude, and situated on flooded plains/alpine rock (some identified as limestone or volcanic). Study based on the reproductive success of whinchats **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** 24 study sites are used, and these are all treated independently. **Intervention description:**	**Primary outcome measures:** Whether birds were carrying prey to feed to chicks **Secondary outcome measures:**	Reproductive success (proportions of territories where juveniles were seen) is lower when greater proportions of the area have been mowed before the date when 80% of the hatched broods would have fledged (based on observations of prey carrying) (r = -0.503, p = 0.024). The relationship between reproductive success of whinchats and density of	**Limitations identified by author:** Probable underestimation of broods killed by mowers due to mowing occurring before hatching/before hatching recognised by surveyors **Limitations identified by review team:**

2009	No information given as to why or how the study areas were chosen. Whinchat selected as considered an indicator species for the evaluation of Alpine management, and is in decline. No evidence given to show that whinchat reproductive success correlated with that of other bird species of the same habitat.	No interventions made by the author, this study describes the result of agricultural practice. Explanatory variable was the proportion of each study area mown by the time that most (80%) broods had hatched/fledged	Reproductive success	whinchats, or density of passerines in general, is complex. Cannot assume that greater bird density is correlated with greater breeding success, those areas with greater density may be acting as population sinks. Whinchats breed later at higher altitudes than at lower altitudes	1. Study areas are treated as respresentative of 'populations' of whinchats, but no evidence is given for why this should be the case and what decided the boundaries/size and whether they can be treated as independent. 2. In addition to the limitation identified by the author, it is likely that the time by which 80% of parents were exhibiting prey carrying would be later if no nests had been destroyed by mowing, which means that reproductive
Aim of study:					**Evidence gaps and recommendations for further research:**

To investigate the effect of mowing date of meadows on whinchat breeding success in alluvial flooded plains and upland meadows					1. No indication is given to how this study may be extrapolated to include other bird species. Further reseach into the relative fledging times would be useful. 2. A baseline study with no mowing would have made the analyses much more reliable
Study design: 2 - correlation study	**Setting:** Flooded lowland and alpine hay meadows in France	**Control / comparison description:**	**Follow-up periods:** None. Studies carried out during 3 consecutive years but at different sites.		
Quality Score: +		**Sample sizes:** 24 study sites	**Methods of analysis:** Correlation		**Sources of funding:** Not supplied
External validity: 2+		**Baseline comparisons:** None **Study sufficiently powered:**			

		The manuscript doesn't demonstrate the independence of the study sites. Power analysis not shown, but likely to be			

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	c
Study Citation	Broyer, J. (2009). Whinchat Saxicola rubetra reproductive success according to hay cutting schedule and meadow passerine density in alluvial and upland meadows in France
Study Design Category	2
Assessed by & when	Kate Fagan, 26th October 2012

1.1 **Are the source population(s) or area(s) well described?**		Comments:

e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	Neither habitats nor biodiversity of the areas were described. Less than half of the study areas were uplands. The only information given that indicated upland habitat was altitude (and all managed as hay meadows)
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	Because of the lack of description it is impossible to tell
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?**		
Was the method of selection well described?	☐-	No method of selection described
Were there any sources of bias?		

Were the inclusion / exclusion criteria explicit and appropriate?		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Not relevant in this case
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	This question isn't relevant in this case since simple correlation is the only analysis used, and only one aspect of this (cutting time/reproductive success) is relevant to the UER question
2.3 Was the contamination acceptably low?	☐NA	Useful part of the study is correlation between cutting time/percentage of land cut versus reproduction success

Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?		
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Study carried out over three different years (i.e. each site was considered only once, but that could have been during the 1st, 2nd or 3rd year) without considering the effect of different years
2.5 Is the setting applicable to the UK?	☐+	Probably, but so few details are given for the sites that it is difficult to be certain
3.1 Were outcome variables/measures reliable?		Comments:

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	The explanatory variable isn't independent.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed?		Forthe purposes of this review, grazing and floristic diversity would

Were all important positive and negative effects assessed?	☐-	have been useful additional measurements
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	
3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	No power analysis given, but appears to have been sufficiently powered (if all study sites can be considered as independent)
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	For an observational correlation study this was fine
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐+	Simple correlation

Were sub-group analyses pre-specified?		
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	p-values given.
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	Some flaws in the study design
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		Insufficient details given

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Year: **Aim of study:**	**Source population:** **Setting:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** Limitations identified by review team: Failure to control for effect of baseline vegetation composition of plots in detailed comparison of species composition and species attributes between treatments in 1991. Baseline vegetation shoudl have been treated as a covariate.

Study design: **Quality Score:** **External validity:**		**Control / comparison description:** **Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Follow-up periods:** **Methods of analysis:**		**Evidence gaps and recommendations for further research:** **Sources of funding:**
Overall score:					

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?**	☐++	Comments:

Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?		

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments:

2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments:
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments:
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Comments:

2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective.	☐++	Comments:

How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?		
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments:
3.3 Were all important outcomes assessed?	☐++	Comments:

Were all important positive and negative effects assessed by the variables/measurements used?		
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:

3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments:

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	**Comments**:

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	**Comments**:
4.3 Were the estimates of effect size given or calculable?	☐+	**Comments**:

4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	**Comments**:
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐+	**Comments**:

[redacted]

5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments:
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments:

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R. **Year:**	**Source population:** **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** **Limitations identified by review team:**

Aim of study:					**Evidence gaps and recommendations for further research:**

Study design: **Quality Score:** **External validity:**	**Setting:**	**Control / comparison description:** **Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Follow-up periods:** **Methods of analysis:**		**Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	2
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	Comments:
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?**		

Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	

2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	
2.5 Is the setting applicable to the UK?	☐+	
3.1 Were outcome variables/measures reliable?		Comments:

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed?		

Were all important positive and negative effects assessed?	☐-	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	
3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐+	

Were sub-group analyses pre-specified?		
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Year: Aim of study: Study design: Quality Score: External validity: Overall score:	Source population: Setting:	Methods of allocation: Intervention description: Control / comparison description: Baseline comparisons: Study sufficiently powered:	Primary outcome measures: Secondary outcome measures: Follow-up periods: Methods of analysis:		Limitations identified by author: Limitations identified by review team: Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any):

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments:

1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments:
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

3.1 How well was the data collection carried out? For example:		Comments:

-Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Not Sure / inadequately reported	

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Clearly described	Comments:
4.2 Is the context clearly described?	☐Clear	Comments:

For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?		
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Reliable	Comments:

5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit?		Comments:

-how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:
5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved?		Comments:

-were negative / discrepant results addressed?	☐ Not sure / not reported	
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not Sure	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Partially relevant	Comments:
5.6 Conclusions For example:		Comments:

-how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	

6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ Appropriately	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ +	Comments:

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: A. Cherrill **Year:** 1995	**Source population:** Improved grassland within the River Tyne catchment **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** NA **Intervention description:** Observational study	**Primary outcome measures:** Proportion of land infested with *Juncus effusus* **Secondary outcome measures:**	Improved grasslands are significantly more infested by *J. effusus* in the uplands than in the marginal uplands or lowlands, but the reason for this isn't clear.It may reflect	**Limitations identified by author:** None **Limitations identified by review team:**

Aim of study: To use stratified sampling to investigate the distribution and extent of infestation of *Juncus effusus* in improved grasslands in the River Tyne catchment, with an ultimate aim of explaining levels of infestation (although the latter isn't really covered by this study)	A land classification devised by the ITE which assigned each 1 km2 to the most appropriate land class was used for this purpose. Squares for surveying were selected by stratified random sampling according to the abundance of the land classes assigned to the squares within the area of interest. Improved grasslands defined as those with over 20% cover of ryegrass.			local factors and the greater availability of rush seed in adjacent upland habitats. Qualitative consideration of management records (as provided by the Farm Business Survey and consisting of the proportions of cut/grazed and permanent/temporary grassland) didn't explain infestation, but was not part of the statistical analysis.	It appears that only improved grassland with over 25% cover of rush was considered
Study design: Randomised observational study		**Control / comparison description:** None **Sample sizes:** 182 squares	**Follow-up periods:** NA **Methods of**		**Evidence gaps and recommendations for further research:**

Quality Score: + **External validity: -**	**Setting:** The catchment of the River Tyne. Approximately half the squares surveyed were defined as lowland, a quarter marginal upland and a	were surveyed **Baseline comparisons:** None **Study sufficiently powered:** No power analysis, but results reported are highly significant.	**analysis:** Non-parametric analysis of variance was used to compare improved grassland and rush cover between different land classes. The student's t-test was used to compare infestation levels between land classes and landscape type		**Sources of funding:** The Natural Environment Research Council and the Economic and Social Research Council

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	b - appraoches to control rushes
Study Citation	Cherrill (1995). Infestation of improved grasslands by *Juncus effusus* L. in the catchment of the River Tyne, Northern England: a field survey
Study Design Category	2
Assessed by & when	Kate Fagan 29/11/12

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Comments:
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Almost all of the catchment area in question was part of the study, and different land classes were selected for study in proportion with their occurrance.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?**		Only improved grassland was studied, but the selection method was well described.

Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Descriptive study only
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	

2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐NA	
2.5 Is the setting applicable to the UK?	☐++	River Tyne catchment, approximately a quarter of the land considered uplands

3.1 Were outcome variables/measures reliable?		

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed?		Rushes only considered if they grew at more than 25% cover. Only

Were all important positive and negative effects assessed?	☐-	imrpoved grasslands considered
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	
3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐NA	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Explanatory variables weren't considered at all, despite further information from the Farm Business Survey data which could have been used analytically
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐+	Very few analyses carried out

Were sub-group analyses pre-specified?		
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NA	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		Very little use for the question under consideration

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Court, I. Barker, D. Cleasby, I. Gibson, M. Smith, J., Straker, C & Thom, T J. (2001) A survey of yellow Wagtails in the Yorkshire Dales National Park in 2000 and a Review of their Historical Population Status. YDNPA, Grassington.
Study Design Category	2
Assessed by & when	CE Pinches, 12th December 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Yes the Yellow Wagtail, its preferred habitat requirements are well described.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes, survey coverage is within 10 areas within the Yorkshire Dales National Park.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Selection is targeted not random, selection was based on coverage of previous one off wagtail surveys and historical sources identifying areas as supporting high populations of yellow wagtail in the past.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NR	Comments: Not relevant
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes – study sought to look at the impact of management practices on bird breeding success and ultimately populations, in particular to look at impact of meadow cutting on fledging.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NR	Comments: Not relevant
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++	Comments: Other factors considered in discussion but study can not quantify their impacts.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Yes observations to determine fledging success of breeding pairs
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	Comments: Not reported but survey intensity presumed to be same across all survey arase.
3.3 Were all important outcomes assessed? Were all important positive and negative	☐+	Comments: Yes

effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: Not applicable
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: Survey is snapshot of breeding success in one year so is highly influenced by weather conditions in that season. However it enables some comparsion with historical surveys so gives indication of population decline.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NA	Comments: Presentation of observational survey data – simple descriptive comparison where this is possible.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	Comments: No, Not Applicable description of survey results only.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: probably given nature of data though some simple t tests may have been usefully applied where comparative survey methology was used between current and past survey.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NR	Comments: No
Section 5: Summary		
5.1 Are the results of the study internally		Comments: Yes, broadly.

valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	**C)**

Study details	Authors	Court, I. Barker, D. Cleasby, I. Gibson, M. Smith, J., Straker, C & Thom, T J. (2001) A survey of yellow Wagtails in the Yorkshire Dales National Park in 2000 and a Review of their Historical Population Status. YDNPA, Grassington.
	Year	2001
	Aim of study	To determine if there had been a significant decline in the number of breeding yellow wagtails in the Yorkshire Dales National Park.
	Study design	3 one off surveys with some comparison with previous historical data 2
	Quality score	+
	External validity	-
Population and setting	Source population	Yellow wagtail populations with Yorkshire Dales National Park
	Eligible population	
	Inclusion and exclusion	10 areas within YDNP were surveyed for yellow wagtails, based on historical surveys, known

	criteria	presence of high numbers of breeding wagtail and availability of experienced volunteer ornithologists.
	Setting	
Methods of allocation to intervention/control	Methods of allocation	Non –random areas selected on basis described above under inclusion/exclusion criteria.
	Intervention description	Hay cutting – through observation alone.
	Control/comparison description	NA
	Sample sizes	NA
	Baseline comparisons	Some previous information available from survey in 1990, 1991 and 1999, plus information gathered on the historical distribution and abundance of yellow wagtails within surveyed areas to determine level of population change.
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Comprehensive walks allowed location and behaviour of yellow wagtails to be noted within survey areas Repeat visits were then made to sites where yellow wagtails were present and nesting behaviour or evidence of young being fed was recorded. Where it was possible to identify the exact location of a nest, habitat type was recorded and if nest was within a hay meadow, the cutting date was recorded and whether fledging occurred before that date. All areas were repeatedly surveyed until there were no further signs of breeding activity.
	Secondary outcome measures	Additional information on presence of yellow wagtail across a wider area obtained from an enclosed upland breeding wader survey which surveyed 3900 fields in 88 1km^2 across the

		Yorkshire Dales National Park.
	Follow-up periods	Over 1 breeding season all areas were repeatedly surveyed until there were no further signs of breeding activity.
	Methods of analysis	No statistics simple descriptive comparison.
Results		Overall there were 16 confirmed pairs of wagtails with a further 9 possible breeding pairs. The no of pairs confirmed fledging young were 5, with an additional 9 pairs probably fledging. Comparison of the current survey results with limited historical information on the distribution of yellow wagtails in the Yorkshire Dales suggests a serious and widespread decline in range and numbers which appears to have accelerated in the past decade (1990s). The results suggest that yellow wagtails are restricted to areas with less intensive farming practice occurs (i.e typically in upper reaches of the valleys). Earlier cutting dates of hay meadows, especially where there is a change from hay to silage is cited as one of the main causes of the long term decline in yellow wagtail populations, especially when the species fidelity to nesting site is factored in. It is suggested that increased stocking levels may also increase the loss of nests to trampling. Nest building must begin at the end of May for fledging to occur before the ESA Tier 1 cutting date of 7th July. If breeding is delayed due to cold or wet weather in spring, it is possible that hay cutting may take place before the young have fledged. This may account for the more recent declines in yellow wagtail populations over the last decade where there has been a succession of cool, wet springs.
Notes	Limitations identified by author	Pre 1990s good quantitative historical information on the distribution and abundance of yellow wagtails in the YDNP was not readily available , much of the information is anecdotal or based only on a small sample areas. This makes accurate determination of population trends difficult.

	Limitations identified by review team	Survey is snapshot of breeding success in one year so is highly influenced by weather conditions in that season.
	Evidence gaps and/pr recommendations for further research	Research is needed into the relationship between delayed breeding and the impact of cutting dates Research is needed to determine the relationship between nesting sites and the importance of unimproved and wet pastures for feeding during the breeding season
	Sources of funding	YDNP

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Crawley, M.J., Johnston, A.E., Silvertown, J., Dodd, M., de Mazancourt, C., Heard, M.S., Henman,D.F. & Edwards, G.R. (2005) Determinants of species richness in the Park Grass experiment. American Naturalist, 165(2), pp. 179–192.
Study Design Category	2
Assessed by & when	CE Pinches, 24th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Not described in detail in this paper but described fully in other published literature and is MG5 grassland
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: MG5 grassland present on site is known to be representative of that type in lowland England.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	Comments: Means by which treatment plots were allocated is not described in this paper but is known to be non-random. The experiment is 150 years old so set up of treatment plots pre-dates modern concepts of good experimental design.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐+	Comments: From 1991 to 2000, 6 randomly located quadrats measuring 50 cm x 25 cm were located within each plot in early June, vegetation was harvested and dry weight per species determined. Sampling prior to this was irregular and comprised samples taken from 36m2 cut areas - % dry weight of each species determined Lime treatments are described by the author as being confounded with spatial location.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Yes treatments were well explained in Appendices, although it is not possible to quantify precisely the amount of NPK supplied via the FYM treatments. Replication is uneven across the treatments.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes management interventions were well described.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Yes plots are >100m2
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐+	Comments: Park Grass plots were subject to aftermath grazing for the first 20 years, thereafter the aftermath was removed by cutting.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes species rich MG5 representative of wider UK species rich lowland meadow resource.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK	☐+	Comments: Broadly though treatments represent historical practice, for example use of ammonium N

practice(s)?		and fish meal.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Yes objective dry weight assessments of species composition. Soil pH was also recorded for each plot.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: No but this particular park grass study looks at composition of end point/biomass of vegetation and soil pH only in relation to treatment..
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes with the exception of the transient plots, which were split in half in 1989 and N applications were stopped on one half. In 1994, all sub plots on plot 13 were halved and manurin discontinued on half the plots.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Yes, 150 years plus.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they	☐+	Comments: Not reported in detail but the author states that the meadow was relatively uniform at

adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?		baseline.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Comments: Replication is uneven. Issues with experimental power for some treatments have been taken into account in the analyses
4.3 Were the estimates of effect size given or calculable?	☐+	Comments: Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Yes
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Yes these findings provide robust indication of the effect of different nutrient regimes on botanical species richness over a long term treatment regime.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and	☐+	Comments: Yes for MG5.

nationally (i.e. habitat, species)?		

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Crawley , M.J., Johnston, A.E., Silvertown, J., Dodd, M., de Mazancourt, C., Heard, M.S., Henman, D.F. and Edwards, G.R.
	Year	2005
	Aim of study	To test for determinants of species richness by examining the PGE data alongside newly collected data for plots from 1991 -2000.
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Lowland neutral grassland MG5
	Eligible population	As above
	Inclusion and exclusion criteria	
	Setting	2.85 ha of neutral grassland resembling NVC type MG5 Rothamstead, Hertfordshire

Methods of allocation to intervention/control	Methods of allocation	The original experiment consisted of large plots to which different fertilizers are applied. In 1903 most plots were halved and the effects of regular liming tested. This was modified in 1965 with the division of most plots into four sub-plots, three of which are limed to maintain pHs of 5, 6 and 7. The fourth sub-plot receives no lime.
	Intervention description	By 1996, 97 different combinations of liming and fertilizer inputs were present. For full details of treatments and experimental layout please refer to Silvertown et al. 2006, p.g 4 http://www.open.ac.uk/science/biosci/personalpages/j.silvertown/pdfs/Silvertown_et_al_2006.pdf **NPK** Various combinations of inorganic fertilisers (P, K, Mg, Na, nitrate-N, ammonium-N and Si) have been tested since the start; **Lime** Since 1903 the effect of lime has been tested. Lime applied every 3rd year Ground chalk applied as necessary to maintain the soil at pH7,6,5 on sub plots a,b,c respectively with sub plot d representing the nil input control. **FYM** Between 1856 -1863 FYM was applied annually to plot 2 in Nov/Dec at a rate of 35t/ha-1 but was discontinued after eight years because, when applied annually to the surface in large amounts, it had adverse effects on the sward. In 1905 FYM treatments were introduced on three plots, it was the applied every four years at a rate of 35 t per ha, supplying 240kg N, 45 kg P and 350kg K. The plots are cut in mid-June and made into hay. For the first 19 years the re-growth was grazed by

		sheep penned on individual plots but since 1875 a second harvest has been cut and removed immediately.
	Control/comparison description	Yes untreated plot 3
	Sample sizes	Unreplicated
	Baseline comparisons	Yes, 1856, uniformity of the sward was assessed in the 5 years prior to treatments being applied.
	Study sufficiently powered	No.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Botanical composition of the plots has been recorded at irregular intervals with some substantial gaps/ Samples taken between 1862 and 1976 were from 36m2 cut areas - % dry weight of each species determined From 1991 to 2000, 6 randomly located quadrats measuring 50 cm x 25 cm were located within each plot in early June, vegetation was harvested and dry weight per species determined. When the 6 quadrats were aggregated this gave a measure of species richness at 0.75m2 for each plot.
	Secondary outcome measures	
	Follow-up periods	150 + years
	Methods of analysis	Crawley et al. 2005 applied a maximal model (including interaction terms and quadratic terms for continuous explanatory variables) was fitted first then the model simplified involved deletion of variables and reduction of factor levels. Explanatory variables are: experimental treatments: categorical variables with two levels in the case of P and K (applied or not); 3 levels for the type of N (none, ammonium sulphate, or sodium

		nitrate); 4 levels for liming; two levels for the transients; two levels for the organics (organics applied or not) and one continuous explanatory variable (application rate of N) with two covariates; total first cut biomass and soil pH. HH
Results		Subsequent impacts Species richness was greatest on plots that had no experimental inputs >40 and lowest in plots were the soil was strongly acidified by the long term input of ammonium sulphate supplying 144 N kg per ha. Species richness declines from the control plots, through plots receiving P alone, sodium nitrate or ammonium sulphate on their own, N and K together (-P), FYM and P together with K. The largest reduction in species richness are associated with adding N and P together and maximum depression of species richness occurs when N is applied as ammonium sulphate. Only N ($p<0.00001$) and P ($P<0.00001$) had significant main effects on species richness. There was no significant interaction between N and P application ($p=0.14$) the effect of adding N and P together was additive and was responsible for the greatest reduction is species richness attributable to nutrients. There was a roughly linear decline in mean species richness with N application rate for both types of N. Modern species numbers vary from 3 to 44 per 200 m2 among the plots According to the multivariate model of species density variation 50 kg N ha^{-1} $year^{-1}$ added as fertilizer reduces species number by about 6.5 species, ammonium N loses 3 more species than would the same rate of N as sodium nitrate (because of the effect on soil pH), using organic manures rather than mineral fertilisers adds two species on average. Crawley showed that the addition of phosphorous reduced species richness, and application of potassium along with phosphorous reduced species richness further, but the biggest negative effects were when N and P were applied together.

		Liming There was no response to relationship between lime treatment and species richness except in plots receiving nitrogen in the form of ammonium sulphate, where species richness increased sharply with increasing pH. Another critical determinant of the species composition of the plot is the N:P ratio.
Notes	Limitations identified by author	Due to age of experiment there was no randomization of treatments and replication is uneven, treatment combinations are missing and lime treatments are confounded with spatial location.
	Limitations identified by review team	Park Grass plots were subject to aftermath grazing for the first 20 years, thereafter the aftermath was removed by cutting. Botanical analysis of the 3 post 1905 FYM plots difficult to describe because two of them also receive fertilisers or fish guano. Only plot 19 is FYM only and a valid comparator.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NERC, BBSRC and Lawes Trust

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Critchley, C.N.R., Chambers, B.J., Fowbert, J.A., Sanderson, R.A., Bhogal, A., & Rose, S.C.
	Year	2002
	Aim of study	To determine the relationship between a range of British lowland grassland plant community types and a standard set of soil variables - To quantify the levels of soil nutrients and other soil properties for broad grassland type and grassland NVC types - To determine the relative importance of these relationships for different grasslands - To evaluate implications for conservation management of lowland grassland types
	Study design	2 (Correlative study)
	Quality score	-
	External validity	-
Population and setting	Source population	Semi-natural grasslands in Environmentally Sensitive Areas ESAs

	Eligible population	With respect to this study, interested in MG3 sites within Pennine Dales ESA.
	Inclusion and exclusion criteria	
	Setting	Lowland England (below the line of enclosure).
Methods of allocation to intervention/control	Methods of allocation	Original ESA monitoring was through either random or stratified random sampling.
	Intervention description	NA
	Control/comparison description	NA
	Sample sizes	63 quadrats from Pennine Dales – botanical data collected from 5 1x 1m quadrats Note on 6 MG3 samples in sample/ Twenty soil cores were collected immediately adjacent to the edge of each plot and from one randomly selected quadrat per field.
	Baseline comparisons	None
	Study sufficiently powered	No not for MG3.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	% cover estimated for all vascular plant species. Soil pH, extractable K and Mg Total N and organic matter content Both Olsen and resin Extractable P
	Secondary outcome measures	
	Follow-up periods	Soil sampling was undertaken at the same time as the botanical recording.

	Methods of analysis	Individual quadrats and plots were classified within the framework of the NVC by generating 50 random pseudo-quadrats from data in Rodwell 1991 and 1992 for each NVC community and sub-community known to occur in grasslands in the ESA sampled . Actual quadrats from the ESAs were then added passively suing cover values and frequencies to Detrended Correspondence Analysis. Distance from NVC community was worked out for each and this data was used to propulate the summary statistics of soil properties for each community. Relationships between plant community types and soil properties was investigated using CCA.
Results		Across the sample, grasslands of high botanical value were generally associated with lower levels of soil extractable P and K. Low levels of soil P and K were a feature of the most botanically valuable unimproved mesotrophic grasslands. The MG3b (Bromus hordeaceus sub community which is normally linked with disturbance and fertiliser and lime applications was associated with higher pH values than the key species rich community MG3b (Briza media)The semi-improved sub-communities MG6c (Trisetum flavescens) and MG6b (Anthoxanthum odoratum) were separated by having higher and lower pHs respectively. Unimproved mesotrophic grasslands also had relatively low ecological amplitude suggesting that they are potentially sensitive to altered soil properties. These grasslands were differentiated from one another at the sub community by their soil pH. Raising soil pH by lime application, or a long term downward drift where pH has been artificially raised in eth pas could change the identity of these sub communities. They will also be vulnerable to soil acidification and increased N availability resulting from atmospheric deposition of sulphuric and nitrogenous compounds. Soil properties were less important in distinguishing the MG3 and MG5 unimproved grasslands from one another, Differences between them are primarily due to variations in altitude and climate. The unimproved MGb Briza media sub community had low P (community mean 8 mgl^{-1} and K

		128mgl$^{-1.}$
Notes	Limitations identified by author	Soil samples collected from outside the quadrat area of the botanical sampling – validation study showed that variation in soil properties at small spatial scales within sites was in most cases markedly less than between sites or grassland types. Using species data in Rodwell to create pseudo quadrats rather than using real data
	Limitations identified by review team	V small number of MG3 sites in sample n =6.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	MAFF

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Critchley, C.N.R., Chambers, B.J., Fowbert, J.A., Sanderson, R.A., Bhogal, A., & Rose, S.C. (2002). Association between lowland grassland plant communities and soil properties. *Biological Conservation, 105*, 199-215.
Study Design Category	2
Assessed by & when	CE Pinches 25th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Yes semi-natural lowland grasslands well described in lowland England context.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐ -	No insufficient sample of MG3 grasslands – only 6.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	As above.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison)		Comments: NA Correlative study

group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Yes/
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Acknowledged but not controlled for.
2.5 Is the setting applicable to the UK?	☐+	yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Yes
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐+	Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	
3.4 Were outcomes relevant?		

Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐NA	

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	No insufficient sample for MG3
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	In relation to soil factors alone.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Yes.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	No as analysis used ordinations
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential	☐-	Overall results valid but insufficient sample size to adequately characterised MG3 and its sub communities.

confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	No inadequate sampling.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Critchley, C. N. R., Fowbert, J. A., Wright, B.	**Source population:** Species-rich mesotrophic hay meadows in the Pennine Dales	**Methods of allocation:** No allocation - management recorded but not stipulated	**Primary outcome measures:** Differences between plant communities in 1987 compared with 2002	Reestablishment of the target species-rich community was associated with late cutting, absence of cattle grazing and and an early close date for spring grazing.	**Limitations identified by author:** Relatively small number of quadrats already close to the MG3 target community, too few for significant relationships with management information to be identified.
Year: 2007	**Eligible population inclusion & exclusion criteria:**	**Intervention description:** No intervention - correltaive/monitoring study	**Secondary outcome measures:**	They found a significant increase (*p* <0.05) in Ellenberg N	**Limitations identified by review team:**

	From the sites used in two previous studies, one where sites were randomly selected and the other where sites were selected by stratified random sampling, all sites where previous surveys had indicated a significant MG3 community or the potential for reversion to that community were investigated			values between 1987 and 2002 in the modified species-rich sample, (grasslands with close similarity to MG3). Furthermore Ellenberg N-values were more likely to increase at higher soil pH (*p* <0.05) and extractable P (*p* <0.01). In contrast, change in species composition of species-rich MG3 meadows over the 15 year time scale studied was found to be associated with lower soil extractable K values (*p*< 0.01).	In general, the results section gives the statistics for the 'change in species composition' significantly explained by different management variables. In cannot be clear from the results whether this change is towards the target MG3 community, or more diverse, or the opposite of either of these things. No ordination diagram is given. From the conclusions it is clear that the 'change' they refer to is positive, but it is not clear from the results or from the

Aim of study: To investigate whether a reversion to traditional management techniques (as defined by the ESA scheme) has lead to the re-establishment of characteristic upland hay meadow communities, and to identify which managament techniques were more or less successful in this					**Evidence gaps and recommendations for further research:**
Study design: Systematic review	**Setting:** Pennine Dales	**Control / comparison description:** N/A - survey of sites with analysis of previous management	**Follow-up periods:** This study was carried out 15 years after the previous one		
Quality Score: 2-		**Sample sizes:** 116 hay meadows, of which 16 species	**Methods of analysis:**		**Sources of funding:** Defra funded

External validity: 2-		**Baseline comparisons:** Taken from results of previous surveys **Study sufficiently powered:** No due to small number of genuinely species rich MG3 sites n =16.	Similarity coefficients, Redundancy Analysis, GLM (ANOVA and ANCOVA)		

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	a/c - effects of grazing management/nutrient applications on floristic diversity
Study Citation	Critchley, C. N. R., Fowbert, J. A. & Wright, B. (2007). Dynamics of species-rich upland hay meadows over 15 years and their relation with agricultural management

	agricultural management practices. Applied Vegetation Science 10:	
Study Design Category	2	
Assessed by & when	Kate Fagan 21/11/12	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Comments:
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	

1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	 ☐++	
[redacted]		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++	Stratified/random sampling

2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Mainly fully justified
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐NA	
2.5 Is the setting applicable to the UK?	☐++	Pennine Dales

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Mainly objective
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?		Those that were not complete were eliminated from the analysis

	☐+	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	
3.6 Was the post-treatment time interval meaningful?		This was a meta-study rather than an experimental approach. Longevity of management probably differed

Was the follow up long enough to assess long-term effects?	☐+	between sites.

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	No not for those sites that close similarity to a species rich MG3 community in 1987 n =16. No power calculation given
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Possibly too many even - with the number of varibles that were tested, the possibility of a type I error is high and this hasn't been corrected for.

4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Stats given
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)?	☐_	No as the sample of true species rich MG3 meadows is too small.

Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐_	Shifts in species composition notably a reduction in number of forb species were associated with lower levels of soil extractable K. This finding was specific to the very small number of species rich upland hay meadows (*n* =16) monitored in the sample reducing the reliability of this evidence.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: C. L. Devereux, C. U. McKeever, T. G. Benton and M. J. Whittingham **Year:** 2004	**Source population:** Lapwings - presumably from broods within the RSPB reserve, though not specifically stated. Starlings locally captured. **Eligible population inclusion & exclusion criteria:** Fifteen colour-winged lapwing chicks aged 5-10 days from 10 broods were selected - no selection	**Methods of allocation:** Starling expt. - 1 sward height treatment was randomly allocated to one of eight patches in a field. 4 cages were arranged in a square within patches. Lapwing expt. - No allocation details given. Experimental design seems to have been determined by another experiment. **Intervention description:** Starling expt. - patches mown to either 3 cm or 13 cm. Trials lasted 15 mins after probing for prey started. Each bird received one replicate of each treatment. Presentation order was randomized. Lapwing expt. - during the year of the expt	**Primary outcome measures:** Starling expt. - number and type of prey captured, the frequency of probes and roots, number of steps walked. Lapwing expt. - 'peck rates' and 'successful peck rates'. Sward height and soil moisture levels were made wherever birds were foraging after they had finished.	Starling expt. - Starlings spent 29.9% more time foraging on short swards, and captured 33.2% more prey on short swards. Both of these results were highly significant. There was no difference in intake rate (ie captures per second of active foraging), indicating that the amount of extra time on short swards was responsible for the extra prey. Prey was largely leatherjackets. Lapwing expt. -	**Limitations identified by author:** Starling expt. - leatherjackets can redistribute when conditions are unfavourable, but the authors considered that the heights involved would not cause this - and this seems to have been upheld **Limitations identified by review team:** The only problem with this study is the lack of fully random (or properly described?) selection of birds or allocation of

Aim of study: To find the effect of grassland sward height and soil moisture on two declining bird species, lapwing and starling **Study design:** Control trial, not apparent whether it is fully randomised or not **Quality Score: +** **External validity: +**	selection information given. 20 locally captured starlings were housed in indoor cages before being used in the enclosures for testing. Again no selection information **Setting:** Lapwing study on agriculturally improved and semi-improved grassland with rig and furrow on Gruinart Flats on the Isle of Islay. Starling study on intensively managed pasture, with 0.5 m3 mesh enclosures, at University Farm, Wytham	during the year of the expt. water levels and fertiliser levels at Gruinart flats were manipulated as part of a wider expt., causing differing conditions of soil moisture and sward height. **Control / comparison description:** NA, as above **Sample sizes:** 15 lapwing chicks observed during 4 hours. Sample size was presumably 15 (or 10??), but this isn't stated (or clear). 20 locally trapped starlings received one replicate of each treatment. **Baseline comparisons:** None **Study sufficiently powered:** No power analysis, but lots of	**Secondary outcome measures:** **Follow-up periods:** None **Methods of analysis:** General linear models and general linear mixed models. Brood (lapwings) and individual (starlings) were entered into the model to control for repeated measures effects	Lapwing expt. Successful peck rates were significantly related to overall peck rates, so on all proceeding analyses only the overall peck rate was used. Foraging rates declined significantly as sward height increased. There was no difference in the number of surface invertebrates found (through pitfall traps) in long and short swards. Lapwing chicks foraged for longer in furrows (short sward) than rigs	allocation of treatments **Evidence gaps and recommendations for further research:** **Sources of funding:** Authors supported by the University of Stirling, RSPB, NERC, BTO and BBSRC

	Wytham, Oxfordshire	significant results, so it appears so	measures effects. Significance of covariance was analysed using the Wald statistic.	sward) than rigs (longer sward) even though the furrows had a lower abundance of food. Soil moisture was not found to be a predictor of foraging behaviour.	

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	a/c
Study Citation	Devereux, C. L., McKeever, C. U., Benton, T. G. & Whittingham, M. J. (2004). The effect of sward height and drainage on Common Starlings (*Sturnus vulgaris*) and Northern Lapwings (*Vanellus vanellus*) foraging in grassland habitats
Study Design Category	2
Assessed by & when	Kate Fagan 03-12-12

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	No plant community information
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	No information on this
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias?	☐NR	Again no information

Were the inclusion / exclusion criteria explicit and appropriate?		

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	No information about how birds were selected (presumably randomly?), but there is an element of randomisation of the patches in the starling experiment
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Well justified
2.3 Was the contamination acceptably low?	☐++	Contamination seems unlikely

Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?		
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++	No confounding factors likely
2.5 Is the setting applicable to the UK?	☐+	Yes, but not necessarily relevant to upland hay meadows
3.1 Were outcome variables/measures reliable?		The outcome measures are fairly subjective, but the perameters have been well documented in the methods

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐+	One of the starling trials was abandoned because no foraging happened within the first 10 minutes
3.3 Were all important outcomes assessed?		

Were all important positive and negative effects assessed?	☐++	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Successful peck rates were found to be correlated with total peck rates, so the latter were used for all analyses
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	
3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐NA	One-off study

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	No power analysis, but lots of significant results so muct have been sufficiently powered
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Analysis not clearly described
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐+	Appeared to be, but not clearly described in all areas

Were sub-group analyses pre-specified?		
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	All test statistics given
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	A few areas where there could be bias
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		Results shouldn't be too habitat-specific

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R. **Year:**	**Source population:** **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** **Limitations identified by review team:**

Aim of study:					**Evidence gaps and recommendations for further research:**

Study design: **Quality Score:** **External validity:**	**Setting:**	**Control / comparison description:** **Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Follow-up periods:** **Methods of analysis:**		**Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	2
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	Comments:
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?**		

Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	

2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	
2.5 Is the setting applicable to the UK?	☐+	
3.1 Were outcome variables/measures reliable?		Comments:

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-		
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR		
3.3 Were all important outcomes assessed?			

Were all important positive and negative effects assessed?	☐-	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	
3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐+	

Were sub-group analyses pre-specified?		
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Year: Aim of study: Study design: Quality Score: External validity: Overall score:	Source population: Setting:	Methods of allocation: Intervention description: Control / comparison description: Baseline comparisons: Study sufficiently powered:	Primary outcome measures: Secondary outcome measures: Follow-up periods: Methods of analysis:		Limitations identified by author: Limitations identified by review team: Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any):

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments:

1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments:
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

3.1 How well was the data collection carried out? For example:		Comments:

-Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Not Sure / inadequately reported	

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Clearly described	Comments:
4.2 Is the context clearly described?	☐Clear	Comments:

For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?		
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Reliable	Comments:

5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit?		Comments:

-how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:
5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved?		Comments:

-were negative / discrepant results addressed?	☐ Not sure / not reported	
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not Sure	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Partially relevant	Comments:
5.6 Conclusions For example:		Comments:

-how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	

6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ Appropriately	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ +	Comments:

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Edwards, A. R and Younger, A. **Year:** 2006 **Aim of study:**	**Source population:** Plant communities from two hay meadow areas in the Pennines, one of 6 meadows and one of 4 meadows. **Setting:** Cumbria/North Yorkshire	**Methods of allocation:** Samples of hay taken randomly from a total of 10 different meadows. Hay fed to animals, leading to manure, kept for up to a year. Seed germination trials carried out at each stage	**Primary outcome measures:** Seed germination from hay, manure of different ages, and in vitro digestion techniques **Secondary outcome measures:**	Seeds from a fairly high proportion of the species found in the meadows were present in hay, but grasses were by far most abundant. Far fewer, mainly grass species, were found in the controlled manure. These were stable up until a manure age of 3 months; thereafter viability dropped off	**Limitations identified by author:** Variation can occur between animals, and only two animals used to produce manure **Limitations identified by review team:**

To investigate whether manure may be beneficial in maintaining or improving plant diversity				quickly. Manure that was not controlled experimentally had a much higher diversity of seed, probably contamination from hay. In vitro digestion maintained viability for most tested species other than Myosotis arvensis	**Evidence gaps and recommendations for further research:** The surmised reason for the lack of species of conservation interest is the early hay cut, before many of the species of interest have set seeds. The authors advise similar consideration of the effects of late hay cuts
Study design: Randomised control trial		**Control / comparison description:** In vitro digestion fo 5 typical uplands hay meadow species kept within manure for the same time periods.	**Follow-up periods:** Seed germination trials lasted a year		
Quality Score: ++ **External validity:** ++		**Sample sizes:** 2x1kg samples at each stage for field-based study. 5 samples throughout in vitro digestion studies **Baseline comparisons:** NA	**Methods of analysis:** GLM		**Sources of funding:** Source of funding not given. One author based at Reading University, the other at Newcastle University

Overall score: 1++		**Study sufficiently powered:** No power analysis, but seems sufficiently powered.			

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	b - methods to maintain floristic diversity
Study Citation	Edwards, A. R. and Younger, A. (2006). The dispersal of traditionally managed hay meadow plants via farmyard maure application
Study Design Category	1 - randomised control trial
Assessed by & when	Kate Fagan, 13th November 2012

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Grid references along with plant communities (Domin)

1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Hay samples taken randomly from hay bales, plant community information taken from four (or six) randomly placed 50 cm x 50 cm quadrats per meadow, avoiding edge effects
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Apparently typical of traditionally managed hay meadows, but method of selection not described other than that

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	All elements of the study were randomised
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Very comprehensive descriptions of methods

2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Yes - all seed germination tests were undertaken for a full year
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	No record of any contamination
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups?	NR/NA	Comments:

Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?		
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Study designed to be representative of the Pennines hay meadows
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Representative of more traditional rather than intensive management

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)?	☐++	Outcome measure objective - seed germination. Sometimes to genus rather than species. Difficult to see any source of bias

Was there any indication that measures had been validated/other QA?		
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	In the in vitro digestion part of the study, very few Myosotis arvensis seeds germinated after they had been left for more than 3 months. Data analysis was adjusted accordingly. All other data complete
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:

3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	 NA	No surrogate measurements
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:

3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	A year is usually considered sufficient for seed germination trials

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	NA	Not an in situ study
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?		Power calculation not given, but study appears sufficiently powered in all areas

A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	
4.3 Were the estimates of effect size given or calculable?	☐++	**Comments**:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for?	☐++	All data analyised using Generalized Linear Models

Were any sub-group analyses pre-specified?		
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐++	All stats given in full, with confidence intervals and p values where appropriate
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	Very well-designed study
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++	Source population within the area of UER interest

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Year: **Aim of study:**	**Source population:** **Setting:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** Limitations identified by review team: Failure to control for effect of baseline vegetation composition of plots in detailed comparison of species composition and species attributes between treatments in 1991. Baseline vegetation shoudl have been treated as a covariate.

Study design: Quality Score: External validity: Overall score:		Control / comparison description: Sample sizes: Baseline comparisons: Study sufficiently powered:	Follow-up periods: Methods of analysis:		Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias?	☐++	Comments:

Were the inclusion / exclusion criteria explicit and appropriate?		

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments:
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments:

2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments:
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Comments:

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments:
3.2 Were all outcome measurements complete?	☐++	Comments:

Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant?		Comments:

If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments:

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	**Comments**:
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	**Comments**:

4.3 Were the estimates of effect size given or calculable?	☐+	**Comments**:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for?	☐+	**Comments**:

Were any sub-group analyses pre-specified?		
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐+	**Comments**:
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments:
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments:

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R. **Year:**	**Source population:** **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** **Limitations identified by review team:**

Aim of study:					**Evidence gaps and recommendations for further research:**
Study design:	**Setting:**	**Control / comparison description:**	**Follow-up periods:**		

Quality Score:		Sample sizes:	Methods of analysis:		Sources of funding:
External validity:		Baseline comparisons: Study sufficiently powered:			

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	2
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?**		Comments:

e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled?		

Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	
2.5 Is the setting applicable to the UK?	☐+	

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	Comments:
3.2 Were all outcome measurements complete?		

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐-	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	

3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	

4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	

5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Year: Aim of study: Study design: Quality Score: External validity: Overall score:	Source population: Setting:	Methods of allocation: Intervention description: Control / comparison description: Baseline comparisons: Study sufficiently powered:	Primary outcome measures: Secondary outcome measures: Follow-up periods: Methods of analysis:		Limitations identified by author: Limitations identified by review team: Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any):

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments:
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions?	☐ Clear	Comments:

-is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?		
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question?		Comments:

-Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	

3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Not Sure / inadequately reported	Comments:

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Clearly described	Comments:
4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments:
4.3 Were the methods reliable? For example: -was data collected by more than one method?	☐ Reliable	Comments:

-is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?		

5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments:
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:

5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	☐ Not sure / not reported	Comments:
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not Sure	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Partially relevant	Comments:

5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	Comments:

6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ Appropriately	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ +	Comments:

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	**What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?**

Study details	Authors	Fuller, R J
	Year	1996
	Aim of study	• To review of the relationships between grazing, principally by sheep and bird populations; • To summarise the recent trends in sheep stocking that have occurred throughout Britain; • To described the potential mechanisms by which grazing may affect upland birds and consider the evidence available for these mechanisms. • To place grazing in context with other factors as possible determinants of changes in bird populations • To outline the ornithological implications of a reduction in grazing • To consider the implications of different grazing intensities for individual species of birds • To identify key areas for future research of the relationship between birds and grazing.
	Study design	3(Review)
	Quality score	+

	External validity	+
Population and setting	Source population	Generalist farmland birds (Breeding season birds) mainly waders, lapwing and songthrush or passerines. In winter foraging waders and passerines.
	Eligible population	British bird populations
	Inclusion and exclusion criteria	This review focuses mainly on open upland habitats . However it also considered certain impacts of grazing on bird use of in-bye land, including hay meadows, and best available evidence on the sward structure preferences for a number of breeding waders.
	Setting	UK - concentrating mainly on open, unenclosed upland habitats.
Methods of allocation to intervention/control	Methods of allocation	NA
	Intervention description	NA review
	Control/comparison description	NA review
	Sample sizes	NA
	Baseline comparisons	NA
	Study sufficiently powered	NA
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	NA
	Secondary outcome measures	NA
	Follow-up periods	NA
	Methods of analysis	NA

Results		During the last two to three decades bird populations, especially of breeding waders, have collapsed in many areas of moorland-edge enclosed pastures. This is thought to be linked with increased grazing pressure by sheep, especially through the associated land improvements including drainage, fertilising and reseeding. Such habitat changes may also have affected moorland birds, many of which use marginal farmland for feeding. . Grazing pressure was found to affect different species in different ways, with precise mechanism being species specific. There is a pronounced dichotomy in the sward height preferences of bird species breeding in grassland. Lapwing benefit from a moderate to high level of grazing, maintaining low but not structurally uniform vegetation. In contrast the other principal breeding birds of meadows snipe *Gallingo gallingo*, redshank *Tringa tetanus*, curlew *Numenius phaeopus*, whinchat Saxicola rubetra and skylark *Alauda arvensis*, prefer lighter grazed, tussocky vegetation
Notes	Limitations identified by author	-
	Limitations identified by review team	-
	Evidence gaps and/pr recommendations for further research	The review found "that exceedingly little is known about the ecological relationships between grazing and upland bird populations" and identifies a clear need for specific research into the implications for bird populations of heavy grazing by sheep and also the implications of reduced grazing – but first some means of accurately tracking upland bird populations is required together with quantitative information on long term population trends. There is also scope for improving our knowledge of now different grazing regimes affect the preferred vegetation structures and patterns of selected bird species – including use of mosaics.
	Sources of funding	JNCC

Quality Assessment Checklist: Qualitative Study v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Fuller, R.J. (1996) Relationships between grazing and birds with particular reference to sheep in the British Uplands. BTO Research Report No 164.
Study Design Category	3
Assessed by & when	C.E. Pinches

Section 1: Theoretical approach		
1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments: Yes, reviews available evidence on on the various mechanisms by which upland sheep grazing effects upland bird populations.
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments: Objectives and parameters clearly set out as are what is and isn't included. The review focuses on the effect of sheep grazing on bird populations on the open, unenclosed uplands of Britain but touches on impacts on in bye land and its usage by upland bird species.
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling	☐ Not Sure / inadequately reported	Comments: It is not clear what approach has been applied to searching the literature for relevant evidence/information. However the number of citations referred to in the text is lengthy and indicates a comprehensive review has taken place.

strategy theoretically justified?		

Section 2: Study Design		
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure / inadequately reported	Comments: It is not clear what approach has been applied to searching the literature for relevant evidence/information. However the number of citations referred to in the text is lengthy and indicates a comprehensive review has taken place.

Section 3: Data Collection		
3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Appropriately ☐ Not Sure / inadequately reported	Comments: Not clear how references were searched for and whether this was systematic.

Section 4:Trustworthiness		
4.1 Is the role of researcher clearly described? For example:	☐ Not	Comments:

-has the relationship between the researchers and intervention group been adequately considered?	described	
4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments: Yes context of declining upland bird populations and increases in the upkand sheep population very well described.
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Not Sure / not reported	Comments

Section 5: Analyses		
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments: No explicit quantitative analysis was conducted, instead the literature is reviewed and reported.
5.2 Is the data 'rich'? For example:	☐ Rich	Comments: A wide diversity of literature has been used.

-how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?		
5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	NA	Comments: NA literature review only
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Convincing	Comments: Findings are clearly presented and well evidenced .
5.5 Are the findings relevant to the aims of the study?	☐ Relevant	Comments: Yes relevant.
5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Clear	Comments: The conclusions are clear and any areas of speculation are acknowledged as are further areas which would benefit from research.

Section 6: Ethics		
6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered? - Was the study approved by an ethics committee?	☐ Not Sure / not reported	Comments: NA

Section 7: Overall Assessment		
As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ ++	Comments: Well conducted – despite there being no description of the method used to search the literature, the list of references evaluated and cited is comprehensive.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: R. E. Green, G. A. Tyler, T. J. Stowe and A. V. Newton **Year:** 1997	**Source population:** Breeding corncrakes on North and South Uist, the Isle of Coll, in the Shannon Callows and an area of County Donegal **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** NA - observation only, no treatments **Intervention description:** NA - as above	**Primary outcome measures:** Breeding success of nesting corncrake females **Secondary outcome measures:** Through the use of a model	It seems that most females will produce two broods, whether or not the first brood was successful. An average of 1.85 broods was produced per female. In the habitats surveyed,	**Limitations identified by author:** None **Limitations identified by review team:** Some of the

Aim of study: To use the results of intensive studies of the breeding biology and success of radio-tagged female corncrakesand the effects of mowing on nest and chick survival in a few study areas to construct a simulation model of nesting and chick-rearing and then explore the effects on breeding success of altering mowing practices	All females trapped in cage traps set in lines of drift fences were eligible, as well as nesting females found later during field observation within the study areas	**Control / comparison description:** NA **Sample sizes:** A total number	produced from the measurement of a number of interim markers of actual breeding success, modelled breeding success under changed perameters **Follow-up periods:** Breeding success measure from incubation until independence of offspring, at which point it was no longer possible to locate them	only 23% of first nests were found in grass meadows, meadows apparently not providing sufficient cover early in the season. 80% of of repeat or replacement clutches were found in habitats liable to be mown. Accoring to the simulation model, inside out mowing results in a substatial increase in corncrake productivity over outside in mowing, especially when the mowing date is early. Moving the mowing date from the end of June to the beginning of September resulted in an extremely large increase in	behaviour of corncrakes that seemed to most affect the simulation model may well be due to the northern situation of the study sites. The model is likely to be quite different in England. **Evidence gaps and recommendations for further research:**

Study design: Non-random observational study used to prduce a simulation model **Quality Score: +** **External validity: +**	**Setting:** North Uist and South Uist (Outer Hebrides, Scotland), the Isle of Coll (Argyll, Scotland), in the Shannon Callows (Republic of Ireland) and Co. Donegal	of 58 nests were used to prepare the model, but not all of these were used in all areas of the model. This depended on the aspect of breeding success considered. Some nests were deleted, if there was a logical reason, to improve the accuracy of the model **Baseline comparisons:** NA **Study sufficiently powered:** No power analysis. Standard errors of the model when used with real data were approximately 17%, which seems acceptable given the number of variables involved and the differences in outcomes	**Methods of analysis:** The rate of egg-laying was estimated by least squares regression. All other measures of breeding success were measured through the use of Kaplan-Maier diagrams and modelled on the basis of normal distributions. A simulation model was produced to encompass all measured aspects of corncrake breeding	productivity, but even using an intermediate date of the model (beginning of August) almost doubled the productivity in most iterations of the model presented.	**Sources of funding:** Royal Society for the Protection of Birds and the Nature Conservancy Council

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	c - effect of cutting date/pattern on breeding success of corncrakes

Study Citation	Green, R. E., Tyler, G. A., Stowe, T. J. And Newton, A. V. (1997). A simulation of the effect of mowing of agricultural grassland on the breeding success of the corncrake (*Crex crex*). Journal of Zoology 243: 81-115
Study Design Category	2
Assessed by & when	Kate Fagan 02-12-12

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	Very little description of the habitat
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?**	☐+	Corncrakes were sampled if caught in the traps set up. No mention of whether traps may cause any bias or whether they will catch birds that are totally representative of the population.

e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?		
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	No detail given of how the areas were selected, or of the hay meadow habitat surveyed (eg NVC community, whether the areas were uplands or not)
[redacted]		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	No treatments

2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Very comprehensive measurements of all aspects of corncrake nesting success
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	No treatments
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Biggest problem was loss of radio-tagging, much worse in some areas than in others, but this was compensated for statistically

2.5 Is the setting applicable to the UK?	☐+	Setting is mainly in the UK, but the habitats may not be similar to English upland hay meadows

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Objective, but most important results obtained through modelling
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?		No, due to problems with radio-tagging, but adjusted statistically

	☐+	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	No treatment

3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Yes, but longer would have been better. Impossible because of problems with radio-tagging and length of vegetation making location of birds too difficult

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	No power calculation, but appears OK
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	

4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NA	No intervention effects
5.1 Are the results of the study internally valid (i.e. unbiased)?		

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Although really not certain that results can be generalised

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R. **Year:** 2012	**Source population:** Whinchat nests in hay meadows **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** This was determined by farmers managing the land as they wanted **Intervention description:**	**Primary outcome measures:** Nest success (survival of chicks to fledging) **Secondary outcome measures:**	Increasing altitude by 100 m delayed laying date by 0.7 days. Models indicate that while early hay cuts lead to low nest survival (10%), there is no significant difference in nest survival between those protected in early-mown fields and those in late mown fields (70%).	**Limitations identified by author:** The success of nest protection doesn't take into consideration the long-term effect of habitat degredation due to more intensive management **Limitations identified by review team:**

	All nests within a certain, pre-described area included in the study. The reasons for chosing the area were not defined	Three groups were defined: those subjected to early cuts (before July 1st); those subjected to late cuts (after July 1st) and those subject to early cuts but where the nests were protected		Nest protection is very expensive, however.	If habitat degredation has occurred due to early mowing, it is possible that those fields that are generally mowed early aren't as suitable for breeding success and nest survival may be linked to this. 'Field effect' hasn't been considered

Aim of study: To investigate whinchat nest success in early mown meadows, late mown meadows, and under protection in early mown meadows					**Evidence gaps and recommendations for further research:**
Study design: Observational quantitative study **Quality Score:** + **External validity:** +	**Setting:** Alpine valley hay meadows	**Control / comparison description:** As above **Sample sizes:** 104 nests were studied in total, 41 in the first category, 26 in the second and 37 in the third **Baseline comparisons:** N/A **Study sufficiently powered:** No power analysis, but yes	**Follow-up periods:** To fledging **Methods of analysis:** A time dependent model of nest survival using the three categories as explanetory variables. Binary logistic regression		**Sources of funding:** Funding source not given. All authors from the Swiss Ornithological Institute

Name of Review Sub-topic (if any): Hay Meadows

Review Question	c - effect of cutting time on whinchat nest survival
Study Citation	Gruebler, M. U., Schuler, H., Horch, P and Spaar, R. (2012): The effectiveness of conservation measures to enhance nest survival in a meadow bird suffering from anthropogenic nest loss
Study Design Category	2
Assessed by & when	Kate Fagan, 4th November 2012

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Few details given, but sufficient for this question

1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	No details given
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	All nests within a defined area included in the study

[redacted]

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++	All nests within pre-defined area
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Yes
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++	Fields noted as under early or late mowing after mowing - i.e. Contamination not possible
2.4 How well were likely confounding factors identified and controlled?		A confounding factors (imperfect detection of nests, probability of nest destruction not linked to mowing) were well-identified and controlled for

Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++	
2.5 Is the setting applicable to the UK?	☐+	Not identical, but seems comparable from the details given

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Outcome measures objective and reliable - survival or failure of nests through specific, pre-defined indications
3.2 Were all outcome measurements complete?		It is possible that some nests were missed - there were some nests identified after fledging that hadn't

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐+	been identified beforehand. Only a small proportion though.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	

3.5 Were there similar follow up times in exposure and comparison groups?	☐+	All treated in the same way, but nests destroyed through mowing would also have had a chance of failing though other reasons. This was corrected for, as much as possible
3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐++	Mowing could have no feasible effect over longer timescales than those measured

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	No power calculation given, but total of over 100 nests studied, which in this study seems perfectly acceptable

4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Good, detailed analysis that overcame potential limitations well
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	All stats given

5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	I don't know enough about how whinchat populations may differ between England and Switzerland, but habitat management is comparable from what is reported in this study

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Year: **Aim of study:**	**Source population:** **Setting:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** Limitations identified by review team: Failure to control for effect of baseline vegetation composition of plots in detailed comparison of species composition and species attributes between treatments in 1991. Baseline vegetation shoudl have been treated as a covariate.

Study design: **Quality Score:** **External validity:**		**Control / comparison description:** **Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Follow-up periods:** **Methods of analysis:**		**Evidence gaps and recommendations for further research:** **Sources of funding:**
Overall score:					

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias?	☐++	Comments:

Were the inclusion / exclusion criteria explicit and appropriate?		

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments:
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments:

2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments:
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Comments:

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments:
3.2 Were all outcome measurements complete?	☐++	Comments:

Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant?		Comments:

If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments:

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	**Comments**:
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	**Comments**:

4.3 Were the estimates of effect size given or calculable?	☐+	**Comments**:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for?	☐+	**Comments**:

Were any sub-group analyses pre-specified?		
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐+	**Comments**:
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments:
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments:

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R. **Year:**	**Source population:** **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** **Limitations identified by review team:**

Aim of study:					**Evidence gaps and recommendations for further research:**
Study design:	**Setting:**	**Control / comparison description:**	**Follow-up periods:**		

Quality Score: **External validity:**		**Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Methods of analysis:**		**Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	2
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?**		Comments:

e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled?		

Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	
2.5 Is the setting applicable to the UK?	☐+	

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	Comments:
3.2 Were all outcome measurements complete?		

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐-	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	

3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	

4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)?	☐-	

Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Year: Aim of study: Study design: Quality Score: External validity: Overall score:	Source population: Setting:	Methods of allocation: Intervention description: Control / comparison description: Baseline comparisons: Study sufficiently powered:	Primary outcome measures: Secondary outcome measures: Follow-up periods: Methods of analysis:		Limitations identified by author: Limitations identified by review team: Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any):

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments:

1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments:
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

3.1 How well was the data collection carried out? For example:		Comments:

-Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Not Sure / inadequately reported	

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Clearly described	Comments:
4.2 Is the context clearly described?	☐Clear	Comments:

For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?		
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Reliable	Comments:
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit?		Comments:

-how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:
5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved?		Comments:

-were negative / discrepant results addressed?	☐ Not sure / not reported	
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not Sure	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Partially relevant	Comments:
5.6 Conclusions For example:		Comments:

-how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	

6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ Appropriately	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ +	Comments:

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Hochberg, H., & Zopf, D. (2011). *Preservation of forage quality and biodiversity by utilization of mountain meadows.* Grassland farming and land management systems in mountainous regions. Proceedings of the 16th Symposium of the European Grassland Federation, Gumpenstein, Austria, 29th-31st August, 2011
Study Design Category	2
Assessed by & when	CE Pinches, 25th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	No description of botanical composition, altitude, soil conditions.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Authors state they are.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	Unclear given scant description, only one site per grassland type so sample small.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐NR	Comments: Unclear, not reported,
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: All treatments described – not clear if aftermath grazing is taking place, it is not reported,
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: Presumed to be.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐-	The confounding influence of cutting regime should be recognised in interpreting these results, as treatment 1 whilst under higher NPK input was subject to 3 cuts annually, treatment 2 under intermediate NPK input had 2 cuts annually whilst treatment 1 under lowest NPK input had one cut alone.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐-	Comments: Unclear, alliances are mountain hay meadows so presumed to be partially relevant.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Yes in agricultural terms

Section 3: Outcomes		
3.1 Were outcome variables/measures		Comments: It is not clear what nature the botanical

reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	assessment took, it looks to be % biomass. No information is presented on frequency of assessments therefore assumption is that botanical assessment was made at end of experiment only.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: Botanical composition only recorded at 2 time periods, start and endpoint.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐NR	Comments: unclear. % biomass for individual species not reported.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐-	Comments: Insufficient sampling of botanical composition.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐-	Comments: Unknown.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐-	Comments: Not reported and insufficient details provided to determine in sampling and replication sufficient.

A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐-	Comments: No
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐NR	Comments: No details provided
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NR	Comments: Not reported.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐-	Comments: Impossible to know as insufficient information presented on experimental design and analysis.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐-	Comments: Difficult to know as above.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rate of application and timing/periodicity of nutrients and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Hochberg, H & Zopf, D
	Year	2011
	Aim of study	To determine yield, forage quality and botanical composition of grassland communities in response to different fertiliser and cutting regimes.
	Study design	2 (Controlled trial but not fully randomised).
	Quality score	-
	External validity	-
Population and setting	Source population	Mountain meadows in Germany
	Eligible population	As above
	Inclusion and exclusion criteria	None reported.

	Setting	Three permanent grassland types - *Trisetetum*, *Geranio-Trisetetum* and *Meo-Festucetum*
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	Treatments were: (1) first cut at silage stage, 3 cuts annually, optimal amount of fertilizer according to the yeild (nitrogen (N), phosphorus (P), potassium (K)) (200 N, 25P and 220K kg/ha at the Trisetum grassland; 130N, 30P and 149K kg/ha at the Geranio-Trisetetum grassland; 150N, 25P and 220K kg/ha at the Meo-festucetum grassland. (2) first cut at hay stage, 2 cuts annually, 60 kg N/ha/yr, P and K on a level of nutritive yield; (3) first cut at beginning of July, without N fertilization but P and K on a level of nutritive yield, except on Meo-Festucetum. i.e 15P 100K kg/ha at Trisetum grassland and 20P 97K kg/ha at Geranio Trisetum
	Control/comparison description	No nil input control
	Sample sizes	Treatments were replicated 4 x at each site.
	Baseline comparisons	Unclear if any comparison made with starting composition at baseline.
	Study sufficiently powered	Unclear though replication sufficient.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Dry matter yield and parameters of forage quality such as contents of crude fibre, crude protein, energy and minerals as well as dry matter digestibility were analysed. Determination of plant species was conducted according to the method of Klapp-Stählin (Voigtländer and Voss, 1979)
	Secondary outcome measures	

	Follow-up periods	18 years
	Methods of analysis	ANOVAs – no details given of model parameters or whether baseline composition was used as a explanatory variable
Results		The proportion of herbs (biomass) declined linearly from treatment 1 to treatment 3 (High to Low input of N) whereas legumes increased *Geranio-Trisetetum*, and especially in *Trisetetum*. In *Meo-Festucetum* legumes were absent and the proportion of herbs increased linearly from treatment 1 to treatment 3. Total species number was stable: *Trisetetum* 24-27 species, *Geranio-Trisetetum* 36-40 species and *Meo-Festucetum* 11-13 species (data not shown). The grassland communities maintained a typical setting according to the botanical composition. In treatment 3 the sward was characterized by a higher level of indicator plants of poor soils. The three investigated treatments maintained the mountain grassland communities in a condition that had an acceptable range of abundance of typical species. Species-rich communities were stable in terms of the number of indicator species
Notes	Limitations identified by author	None
	Limitations identified by review team	Insufficient information provided on the nature of botanical change other than at the grass, herb, legume level. No information is provided on condition/composition at baseline. ANOVA appears to have been conducted to determine difference in treatment effects not change through time. The confounding influence of cutting regime should be recognised in interpreting these results, as treatment 1 whilst under higher NPK input was subject to 3 cuts annually, treatment 2 under intermediate NPK input had 2 cuts annually whilst treatment 1 under lowest NPK input had one

		cut alone. Limitation to this project is that system examined appears to be under cutting management only, i.e no grazing.
	Evidence gaps and/pr recommendations for further research	-
	Sources of funding	

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Honsova, D., Hejcman, M., Klaudisova, M., Pavlu, V., Kocourkova, D. & Hakl, J.
	Year	2007
	Aim of study	To describe plant species composition and species richness after 40 years of fertilizer application. How is plant species richness, sward height, biomass yield and plant species composition affected by long term applications of N,P and K fertilisers.
	Study design	1
	Quality score	+
	External validity	-
Population and setting	Source population	Alluvial meadow in Czech republic.
	Eligible population	

	Inclusion and exclusion criteria	
	Setting	Cernikovice, 25km south of Prague at 363m in a flat alluvial meadow. Soil type was fluvisol/gelysol with a loamy texture and pH of 6.0
Methods of allocation to intervention/control	Methods of allocation	Randomised block design with 4 blocks, with plots of 5 x6m within them on which 6 treatments were applied. Experimental area was limed occasionally when a decrease in soil pH was detected.
	Intervention description	Mown 3 x a year from 1966 to 1985 and since the late 1980s twice yearly. Unfertilized control PK, N100PK, N200PK, N300PK, N400PK. The two latter treatments were added in 1975. In 1990 the dosages were reduced by half , the following treatments were then applied control, pk, N50PK, N100PK, N150PK and N200PK. N.B this study focuses on the impact of these treatments although there will be a legacy of the past higher inputs. Fertilisers took the form of saltpetre ammonium with lime (27.5% N, 10% Ca), superphosphate (8.5% P, 20% Ca, 10% S) and potash (50%K, 47%Cl). In each year nitrogen fertiliser was applied in April and potassium and phosphate fertiliser were applied in October.
	Control/comparison description	Yes unfertilised control – nil inputs, plus PK only.
	Sample sizes	4 X replication of 6 treatments.
	Baseline comparisons	Yes, at the start of the experiment the alluvial meadows was dominated by grasses with a total cover of 68%, mainly Alopecurus pratensis 17%, Poa pratensis 11%, Festuca pratensis 10% and Holcus lanatus 7%. The cover of legumes was 11% (Trifolium hybridum, T.repens and T.pratense) and that of other dicotyledonous species was 16% (with most abundant species Ranunculus repens,R.acris and Taraxacum species.).

	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Cover of all vascular plant species was visually estimated directly in % in each plot in 2 x 1 by 1m2, before the first harvest in mid May. Actual sward height (ASH) (compressed).
	Secondary outcome measures	Cover according to Functional groups, legumes, herbs and grasses were recognised and further categorisation was made according to mean species height – short gaminoids, short herbs,tall graminoids and tall herbs. Mean height of specie present (obtained from local flora) was weighted according to the total for that species in the quadrats – Potential sward height (PSHPto denote difference between actual (ASH) and potential height. If the difference between the two is positive it means that plants are generally taller than their mean height. Dry matter biomass yield +sum of 1st and 2nd harvest.
	Follow-up periods	Only composition at experimental end point recorded in 2005.
	Methods of analysis	Redundancy Analysis was used to evaluate multivariate plant cover. One way ANOVA was applied to functional groups, species richness ASH and PSH Relationship between ASH and plant species richness, moss cover was evaluated by regression.
Results		Treatment was the most significant predictor of sward structure in the experiment, explaining 32% of the variability in plant cover. The main divergence in the RDA was between treatments without N (control and PK) and those with NPK. If treatments with N were examined separately then the main divergence was between N50 and N100. The effect of treatment on species richness was significant . Species richness varied from 8 to 24

		species per 1m2 plot and was highest in the control followed by the PK treatment and lowest in the N200PK treatment which differed significantly from the other treatments. There was a sig decrease in species richness with increase in sward height ($p<0.001$). With the exception of short graminoids, treatment significantly affected the total cover of all functional groups in 2005. Grasses had lowest cover in the PK and control treatments , which significantly differed from treatments with N. In contrast, tall grasses (A. pratensis and Poa pratensis) had lowest cover recorded in the control and PK treatment and prevailed in all treatments with N. Herbs attained highest cover in the control followed by the PK treatment. This accords with results from other long term NPK experiments. Herbs, tend to posses a lower competitive ability compared to grasses under high NPK. Both these treatments differed significantly from all treatments with N. Tall herbs were recorded in all treatments but the highest cover was found in the control which sig. Diff from all other treatments. . Cover of rhizomatous grasses exceeding 80% was recorded in the N200PK treatment. Legumes were not detected in the N200PK treatment and achieved highest cover in the control and PK treatment. Although there was significant positive effect of the latter treatment, this is counter to other studies and may suggest that P and K concentrations at the study site were not limiting prior to inputs. Decline of legumes in plots receiving high N was caused indirectly by competition with tall grasses and directly due to their high sensitivity to the increased nitrate concentrations in the soil affecting the transport of assimilates from the leaves to underground organs. More generally the absence of a negative effect of PK treatment on species richness is likely to be because the limiting nutrient to biomass production is N at the study site. The only legumes able to tolerate moderate applications of N were Latyhrus pratensis and

		Trifolium repens. It is reported that under the previously higher rates of NPK imposed in the 1966 to 1990, that cover of rhizomatous grasses increased from 66% to 80 -98% in the first two seasons of the N100PK and N200PK treatments. The author the previous study concluded that initial changes in plant species composition caused by fertiliser application ceased within 3 – 6 years.
Notes	Limitations identified by author	-
	Limitations identified by review team	The study reported here makes use of existing fertilizer plots set up in 1966, on which rates of HPK application were double that of those latterly applied since 1990s. There is no recognition in the results section of the legacy that these high input treatments will have had on botanical composition at experimental end point 2005. Changes in sward composition as a response to these earlier treatments are instead reported in the discussion, it is unclear why two analyses could not have been run to examine change through time from 1966 – 1990s and then from 1990s to present day. It would then be easier to determine the relative impact of these treatments.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Not reported.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Honsova, D., Hejcman, M., Klaudisova, M., Pavlu, V., Kocourkova, D. & Hakl, J. (2007). Species composition of an alluvial meadow after 40 years of applying nitrogen, phosphorous and potassium fertiliser. Preslia, 79: 245-258.
Study Design Category	1
Assessed by & when	CE Pinches, 24th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Yes well described
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Unclear but characteristic of relatively species rich alluvial flood plains similar to MG4
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	Not reported.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Fully randomised control trial
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: yes management interventions were well described, inputs were high.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐-	Comments: Yes, but level of input halved midway through experiment. At start in 1966 inputs rates of N were twice as high, they were then halved from the 1990s to the present day – the impact on botanical composition over 1966 – 1990 and 1990 – 2005 is not fully assessed.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: Presumed to be.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes, two cuts applied, details of dates well described. It is not clear whether any aftermath grazing took place and to what extent this was controlled for.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Looks to be fairly typical of MG4 flood plain meadows.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Yes reflects agricultural practice.

Section 3: Outcomes		
3.1 Were outcome variables/measures		Comments: Subjective botanical assessments made

reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	only at end 2005 of study, no indication provided of change, fluctuation in trends over time.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐-	Comments: Botanical composition only recorded at endpoint.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Broadly ok, insufficient detail on shift in community.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Study looks at actual sward height relative to potential sward height, the latter reflects the typical sward height expected given known height of species constituents in sward. The difference between the two is used as a proxy for fertility but could instead be response to water stress (flooding/drought).
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes, but see above sig change in application rates half way through.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐-	Comments: Unknown no baseline data presented.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐++	Comments: Yes sampling and replication sufficient.

A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐+	Comments: Yes, but partially.
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	Comments:Yes
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Partially.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Although the original experimental design is good a major flaw to this study is that insufficient attempts are made to distinguish between two different exposure periods on the same sample plots. The historic very high input treatments implemented between 1966 and early 1990s are likely to have left a legacy through to 2005. It is not appropriate to assign significant treatments effects apparent in 2005 down to the lower input treatments implemented post 1990.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐-	Comments: Findings are externally valid but need to be seen as partially applicable to UHMs being on floodplain grassland in lowland setting and without any aftermath grazing.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Results	Notes
Authors: Jean-Yves Humbert, Jeronme Pellet, Pierrick Buri, Raphael Arlettaz	**Source population:** Mown semi-natural grasslands	**Methods of allocation:** NA - meta-analysis of independent studies	**Primary outcome measures: Comparison of s**pecies richness/abundance from early/late cut sites, qualified by the standard deviation. The following information was used to try and explain variation: sample size, study duration, plot size, ordinal days of the early cut/delayed cut, meadow type	No overall significant effect of delaying the cut was found on plant species richness, but this result was confused by the	**author:** 1. Great heterogeneity in plant species richness, indicating that factors other than mowing time are important - such as fertiliser application, autumn grazing, seed sowing etc. These other factors could not be investigated due to the highly unbalanced distribution of data points in different categories. 2. Large differences in experimental frameworks probably hindered the detection of effects
Year: 2012	**Eligible Population:**	**Intervention description:** NA - meta-analysis of independent studies	**Secondary outcome measures:**		**Limitations identified by review team:** Data points from the same study can't be totally independent, as they were considered to be in analysis

Aim of study: To find the effect of delayed mowing on plants and invertebrates using a meta-analysis	**Inclusion & exclusion criteria:** Included were studies of semi-natural grasslands that are mown annually where the first mowing date was delayed for the purpose of the study and where there was a relevant control mown at an earlier date. The outcome of the studies was species richness/abundance of any taxa. Non-European sites were excluded.	**Control / comparison description:** Only studies that had a comparison with similar meadows or plots that were first mown on an earlier date (control) were used. Studies were only included if treatment and control plots were similar in all management respects, except the date of the first cut, and were located in the same habitat type.	**Follow-up periods:** Studies lasted for between 2 and 40 years	inclusion of a number of studies where the 'early' cut was in July/August. Further investigation found the date of the early cut had a significant negative effect on plant species richness (ie the earlier the cut the more negative the effect of delaying the cut). Heterogeneity was significant in most of the analyses carried out.	**Evidence gaps and/pr recommendations for further research:**
Study design: 3-meta-analysis of both experimental and observational studies	**Setting:** Europe, but most studies were UK-based.	**Sample sizes: 46 independent data points (not necessarily from different studies),** 35 of which were concerned with plants	**Methods of analysis:** Univariate and multivariate random- and mixed-effect models assessed the effect of all the measured variables on the diversity		**Sources of funding:** Funding source not given, but all authors employed by the Institute of Ecology and Evolution at the University of Bern, Switzerland
Quality Score: -		**Baseline comparisons:** See control description above			
External validity: -		**Study sufficiently powered:** No power analysis given, but due to the extremely diverse set of studies and the relatively small numbers involved, it is likely that significant changes went undetected.			
Question	c - effect of cutting dates on floristic diversity				

Citation	Humbert, J.-Y., Pellet, J., Buri, P., Arlettaz, R. (2012). Does delaying the first mowing date benefit biodiversity in meadowland? Environmental Evidence 1:9				
Study category	2				
Assessed by	Kate Fagan 24/11/12				
Section 1:					
1.1 Is the source population or source area well	□+	Hay meadows in Europe			
e.g. Was the country, habitat and biodiversity of the area well described.					
1.2 Is the eligible population or area representative of the source population or area?	□NA	The representitiveness of individual studies was not commented on			
diversity representative of the habitat?					
groups under-represented?					
1.3 Do the selected habitats/flora/fauna or area represent the eligible population or area?	□NA	The representitiveness of individual studies was not commented on			

Was the method of selection well described?					
Were there any sources of bias?					
exclusion criteria explicit and appropriate?					
Section 2: method of allocation to intervention(or comparison)					
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	□++	Method of selecting studies extremely clear			
2.2 Was the selection of explanatory variables based on a sound theoretical	□NR	The variables seem relevant, but the reasons for selecting them wasn't discussed			
2.3 Was the contamination acceptably low?	□NR	Contamination of the individual studies wasn't discussed			
comparison group receive the exposure? If so, was it sufficient to cause important bias?					
2.4 How well were likely confounding factors identified and controlled?	□-	Likely confounding factors were identified but not adequately controlled			

be other confounding factors not considered or appropriately adjusted for?					
Was this sufficient to cause bias?					
2.5 Is the setting applicable to the	++	Mainly UK studies, the rest European			
Section 3:					
3.1 Were outcome measures and procedures	++				
measure subjective or objective.					
outcome measures (e.g. inter- or intra-rater reliability scores)?					
Was there any indication that measures had been validated?					
3.2 Were all outcome measurements	++				
Were all/most of the study population that met the defined study outcome definitions likely to have been identified?					
3.3 Were all important outcomes	++				
Were all important positive and negative effects assessed?					

3.4 Were outcomes relevant?	NA	No surrogate measures			
Where surrogate outcome measures were used, did they measure what they set out to measure?					
3.5 Were there similar follow up times in exposure and comparison groups?	NR	The methods say that studies were only included if treatment and control plots were similar in all management respects, but follow up time wasn't mentioned.			
3.6 Was the follow up time meaningful?	+	Follow up time vastly different in different studies, but at least 2 years			
long enough to assess long-term benefits / harms?					
Section 4: Analyses					
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)?	-	No power analysis given, but the studies were so different from one another that it is feasible that this was the reason for a lack of an overall significant relationship. Differences between studies weren't in the main corrected for.			
A power of 0.8 is the conventionally accepted standard.					

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?					
4.2 Were multiple explanatory variables considered in the	++				
Were sufficient explanatory variables considered in the analysis?					
4.3 Were the analytical methods appropriate?	+	It seems the data were dealt with in the best way possible, but there were difficulties with the study			
Were important differences in follow-up time and likely confounders adjusted for?					
analyses pre-specified?					
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	++	Stats given			
intervals and or p-values for the effect estimates given or calculable?					
Section 5: Summary					

5.1 Are the results of the study internally valid (i.e.	□-	Major differences between the studies used may have hidden meaningful results			
study minimise sources of bias (i.e. adjusting for potential confounders)?					
Were there significant flaws in the study design					
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?	□+				
details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?					

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Jeangros, B. Sahli, A., & Jacot, P.
	Year	2003
	Aim of study	To detect if the effects of an organic fertilization on a permanent meadow are similar to those of mineral fertilizer.
	Study design	2
	Quality score	-
	External validity	-
Population and setting	Source population	N.B. I have only included findings from one of the two grasslands studied, a less intensively managed meadow at Vuissens, with species richness of 41 species. The other is high intensity dairy grassland.
	Eligible population	Meadows in the Jura– unknown botanical composition,
	Inclusion and exclusion	

	criteria	
	Setting	Jura
Methods of allocation to intervention/control	Methods of allocation	No information provided
	Intervention description	4 Treatments – 1. Ammoniumnitrate 100kg/ha-1, Triple P, 110kg/ha-1, K 125kg/ha-1 2. Manure – 15 tonnes/ha-1 3. Manure 15 tonnes/ha-1 + rock phosphate 135 kg/ha-1 4. Manure 15 tonnes/ha-1 plus calcified seaweed 2.78 t/ha-1 Hay cut between 22^{nd} and 26^{th} June, with a second cut made 9 weeks later,
	Control/comparison description	No null input control provided.
	Sample sizes	Vegetation composition recorded at 2m intervals along a25m transect within each treatment plot.
	Baseline comparisons	Yes at the start of the experiment in 1993.
	Study sufficiently powered	Appears to be 2 x replication
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Botanical composition recorded in mid may 1993 and 1998. Before each cut dry matter yield measures were made from 6 x 0.5m samples per plot.
	Secondary outcome measures	-
	Follow-up periods	6 years
	Methods of analysis	No details provided.

Results		Mainly yield focused. Botanical composition was little effected by type of fertiliser and more influenced by site, year or of the period of harvest. There was a significant change in the botanical composition between 1993 and 1998. Plants other than legumes and grasses significantly decreased whilst grasses increased by 10% in all treatments except manure only. Dandelion decreased significantly whilst Cow Parsley increased significantly. White clover seemed to prefer organic manure than inorganic fertiliser, as did Poa trivialis and Dactylis glomerata.
Notes	Limitations identified by author	-
	Limitations identified by review team	Very little reported on methods and analysis – paper is in French , colleague has part translated. No details provided on starting composition with respect to type of grassland of sward other than it is relatively species rich 41 species. Analysis appears crude and significance of botanical changes is unknown. Not clear whether aftermath grazing has taken place.
	Evidence gaps and/pr recommendations for further research	-
	Sources of funding	Not reported.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Jeangros, B., Sahli, A., & Jacot, P. (2003). Are the effects of an organic fertilization on a permanent meadow similar to those of a mineral fertilization? *Revue Suisse d'Agriculture, 35*, 155-160.
Study Design Category	2
Assessed by & when	CE Pinches, 24th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Simply characterised.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐-	Comments: Manner by which experimental sites were selected is not described, insufficient information provided.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	Comments: No information is provided on how plot location was determined (e.g. randomly) within each meadow sampled.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐-	Comments: Unclear, not reported,
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: All 4 treatments well descried and repeatable.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: Presumed to be.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes, two cuts applied, details of dates well described. It is not clear whether any aftermath grazing took place and to what extent this was controlled for.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐-	Comments: Unclear, look to be fairly typical mesotrophic meadows but insufficient data to be sure.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Yes, although addition of Phosphate and calcified seaweed with manure unusual.

Section 3: Outcomes		
3.1 Were outcome variables/measures		Comments: Subjective botanical assessments

reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: Botanical composition only recorded at 2 time periods, start and endpoint.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Broadly ok, insufficient detail on shift in community, results reported unclearly (although this may be translation issue!)
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Insufficient sampling of botanical composition.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Unknown.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐-	Comments: Not reported and insufficient details provided to determine in sampling and replication sufficient.

A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐-	Comments: No
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐NR	Comments: No details provided
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NR	Comments: Not reported.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐-	Comments: Impossible to know as insufficient information presented on experimental design and analysis.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐-	Comments: Difficult to know as above.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Upland Hay Meadows
Review Question	

Study details	Authors	Kirkham et al
	Year	In prep
	Aim of study	To determine vegetation and soil microbial responses to fertilizers and lime applied in a 12 year study at species rich upland and lowland mesotrophic hay meadows
	Study design	1
	Quality score	1+ for FYM and NPK inputs due to equivalence issue, 1++ for lime.
	External validity	EV+
Population and setting	Source population	Study looks at two meadows an MG3b (Upland) species rich sub community and MG5a/c (Lowland) meadow community
	Eligible population	Upland MG3 hay meadows and Lowland MG5 meadows
	Inclusion and exclusion criteria	
	Setting	Upland Hay Meadow at (Raisbeck), Cumbria; Lowland Meadow (Pentwyn) in Monmouthshire, Wales

Methods of allocation to intervention/control	Methods of allocation	Fully randomised block design. Three blocks at each site. Treatments applied to 7m x 5m plots within blocks.
	Intervention description	Total of 15 treatments. 1. Untreated control , 2. Limed (in 2005) control, 3. FYM at 6 t ha-1 annually , 4. FYM at 12 t ha-1 annually, 5. FYM at 24 t ha-1 annually, 6. FYM at 6 t ha-1 triennially 7. FYM at 12 t ha-1 triennially , 8. FYM at 24 t ha-1 triennially, 9. Inorg. equivalent to Tr. 4, 10.Inorg. equivalent. to Tr. 5, 11.Inorg. equivalent to Tr. 7, 12.Inorg. equivalent to Tr. 8, 13.Lime in years 1 (and 7), 14.Lime as Tr. 13 + FYM as Tr 4, 15. Lime as tr. 13 + FYM as Tr 7
	Control/comparison description	Yes 1: Untreated control. Also Treatment reflects continuation of past fertilisers at, LM site this was identical to Tmt 1, and at UHM site this is Tmt 4. In 2005 limed but unfertilised controls were added - Tmt 2.
	Sample sizes	Botanical: 3 Randomly positioned 1m2 quadrats within each plot. Soils nutrients and biochemistry: Random sampling of 5 cores from each treatment plot 5 equally spaced times intervals. Soil microbial assessments: At 5 sampling positions in plots at 5 equally spaced times intervals
	Baseline comparisons	Yes, botanical and soil chemical property baselines set in 1999.
	Study sufficiently powered	No power analysis presented but there is sufficient replication of treatments and sampling.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Botanical - % cover of each species, converted to % of total live cover. Soil - Organic carbon (C); total nitrogen (N), Olsen extractable P, Exchangeable potassium (K) , Magnesium (Mg), Calcium (Ca), Sodium (Na) and pH. Plus PLFA analyses to assess microbial community.
	Secondary outcome measures	Derived vegetation variables; Total species richness; no per m2 of positive indicator species for community; aggregate cover of positive indicators as % of total veg cover; no/m2 of negative indicators; aggregate cover of negative indicators as % of total veg

		cover; Weighted Ellenberg N score (fertility)
	Follow-up periods	12 years
	Methods of analysis	**Botanical data:** ANOVAs with year as a repeated measure; Separate ANOVAs for each variable within each site. Form x Rate x Frequency (FRF) tested the effects of form of fertiliser (i.e FYM v inorganic), the rate of application and the frequency of application and interactions between these factors Lime x FYM frequency (LFF) model tested the effects of the liming regime **Additional mixed modelling** conducted to identify an effect of form, rate or frequency of application on vegetation composition at each site not simply attributable to the mean per year amount of fertiliser applied. Main effects and interactions were explored. 0rdination of vegetation composition in treatments at end point.
Results		**Overall impact on species richness** . The rate of application of fertilisers on both species richness and the number of positive indicator species were shown to be entirely a function of the mean amounts of nutrients per year. The same was true for frequency of application. An earlier analysis of 2005 data from these sites suggested a positive benefit of triennial application compared to annual application at correspondingly lower amounts (Kirkham et al. 2008). However this affect was not seen over longer timescale despite 3 full 3 year cycles being included, and it was thought that the affect picked up in 2005 may have been due to the fact that the last input on the triennial treatments being applied only a month before botanical assessment so treatment effect would have been limited. It was recognised that more subtle effects of application frequency at the plant community of individual species level may not have been picked up by the composite variables analyse. **Effect of form** This study shows no evidence to suggest that inorganic fertilizers supplying equivalent amounts of N,P and K to ecologically sustainable levels of FYM could not be substituted for these FYM treatments at either site. However N.B.

		inorganic fertilizer treatments received substantially less P and less N than the comparative FYM treatments and the impact of these differences may have not been fully accounted for in the analysis in terms of cumulative effects . **Species richness:** Upland Hay Meadow .Treatments delivering up to about 10kg P ha-1 per year (Equivalent to 11 -12 t FYM ha-1 per year) were consistent with maintenance of species diversity within similar MG3 meadows with some indication that lower rates may be beneficial. The low rate annual (6 t FYM ha -1 per year) and medium rate triennial (12 t FYM ha-1 every 3 years) and the organic equivalent to the latter were all significantly more species rich averaged across all years ($p<0.05$) than annual FYM at the medium rate (12 t FYM ha-1 per year) under either liming regime. Botanical quality, in terms of increased no and cover of positive indicator species and decline in negative indicator species cover occurred in fertiliser treatments incorporating lower nutrient inputs than the historic management (namely <12 t ha-1 annually). Lowland meadow Only about 3 kg P ha-1 per year applied as FYM (equivalent to about 4 t FYM ha-1 per year) or 5-6 kg P ha-1 pear year as inorganic fertilizer were sustainable at the lowland MG5 site. **Plant community composition:** Most treatments at the Upland Meadow site retained a close affinity to MG3b, with the exception of high rate of annual FYM which had moved to a position intermediate between MG3a and MG3b, MG3a representing the less species rich sub community. **Response to lime**: Liming alone had little effect or no detrimental effect on vegetation at either site, but reduced botanical quality of vegetation when applied in conjunction with annual FYM at the lowland meadow site. Occasional liming to raise soil pH to 6.0 appears to be consistent with maintaining vegetation quality within MG3 plant communities. **Soil microbial community:** Fungal to bacterial ratio showed no response to treatment suggesting that low amounts of nutrients are added to the soil by fertilisers but also that there is high resistance of fungal dominated communities of species rich grasslands

		to environmental change **It was considered that responses of vegetation may in part reflect historic adaptations to nutrient and lime inputs.**
Notes	Limitations identified by author	There was a recognised lack of equivalence in nutrient supplied by FYM treatments compared with inorganic comparators between 1999 to 2006. The inorganic equivalent treatments are considered to have supplied substantially less P in particular than was estimated for their FYM counterparts. Composite botanical variables analysed may have masked subtle species specific responses to differences in frequency of application There was a general decline in species richness across all treatments including the control at the lowland meadow site, the authors suggest that a series of late hay harvests may be the cause.
	Limitations identified by review team	Cumulative effect of N and P supplied in FYM may have been underestimated.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Defra, CCW and EN/NE

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	4.3.4.a What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Kirkham et al . (In prep) Ecologically sustainable fertility management for the maintenance of species rich hay meadows: a 12 year fertiliser and lime experiment.
Study Design Category	1
Assessed by & when	CE Pinches, 26th September

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Comments: Plant communities, annual rainfall, soil chemical properties and historic management all well described and characterised in methods and more contextual detail given in discussion.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Manner by which experimental sites were selected is not described but is not random. Sites chosen to be representative of the MG3b (Upland) species rich sub community and MG5a/c (Lowland) meadow communities. Discussion considers how representative study sites are of NVC community generally: SoilpH was slightly low at both sites in 1999, 5.18 and 5.01 at the upland and lowland meadows respectively Olsen extractable P was within range typical of MG3 grasslands, but at lowland meadow the value was considerably lower than normal range of MG5. Soil K levels were high than average at both sites.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	Comments: No information is provided on how plot location was determined (e.g. randomly) within each meadow sampled.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: The experiment employs a fully randomised block design, with three replicate blocks at each site.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: All 16 treatments, including an untreated control are well described and would enable replication. The comparison is appropriate and is an untreated control (no lime/no FYM/no fert). and at each site one treatment reflects continuation of past treatment providing a no change control
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: The ADAS MANNER model was used to predict the amount of plant available N supplied by the FYM, between 1999 and 2006. From 2007 onwards a refined model was used. It is not completely clear from the paper what to what extent N was under estimated. Similarly, the assumed availability of P was increased by 20% (from 60 to 80%) to reflect new research from 2007 onwards. The inorganic equivalents rates were altered accordingly to reflect changes in nutrient supply from the FYM and one set of statistical analyses specifically sought to control for this issue, but impacts of this discrepancy may not have been fully accounted for in the results. Due to the Foot and Mouth disease in 2001 no treatments were applied. This meant that treatments requiring annual applications only received 11/12 (92%) of the intended amounts).
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: In 2005 a decision was taken to lime the untreated control plot in order to ensure pH was reflective of the agricultural norm (pH 6). For the lowland meadow site, this still left one untreated plot within each block reflecting past management. For the UHM site, new untreated control plots were established in random locations adjacent to each block. In addition liming occurred on all fertiliser plots post 2005 if they declined to a pH of 5.5. These changes were well documented and accounted for so are unlikely to cause bias. Whilst plots sizes were small, treatments were applied by hand to each plot to minimise contamination between plots.
2.5 Were any other other intervention(s)	☐++	Comments: Yes, a traditional hay cutting and grazing

received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?		regime were applied across all treatments. The details of this are well described.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes both meadow types are representative of unimproved meadow communities in the lowlands and uplands respectively.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Yes, the range of nutrient rates applied reflect rates historically allowed within agri-environment schemes guidelines together with lower rates

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Both - Subjective botanical assessments - % cover of each species present. Objective - soil sampling for soil nutrients and measures of microbial community structure. Quadrats initially randomly positioned within plots but then fixed. Cover estimates for each species converted to % of live veg cover in each year to reduce year to year variation.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Due to the Foot and Mouth disease in 2001 no botanical assessments were carried out at the Upland site.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes botanical and soil measures are appropriate.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Yes, derived and composite variables appropriate but may be useful to have looked at individual species responses too.
3.5 Were there similar post-treatment time intervals in exposure and comparison	☐+	Comments: Broadly yes, but due to the addition of lime to the original control at the Upland Hay Meadow

groups?		site, a new non limed non fertilised control was set up in 2005 six years into the experiment. As a consequence this control was not included in some of the analyses as it would have made the models unbalanced.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Probably, this experiment is unusual in running for 12 years but even so the full impact of the management treatments may not have become apparent within this timescale.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: Yes, analysis tested for differences over time by using year as an explanatory variable in repeated measures ANOVA. Block was also explicitly used as a explanatory variable or Random factor in the mixed modelling models so any variance attributable to blocking could be determined.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	Comments: No power analysis conducted but there is suitable replication of treatments and the sampling within these treatments is adequate.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Yes, only treatments which were fully factorial were included with the ANOVA models. This meant that only the limed control was included. Similarly adjustments were made to the mean amounts of nutrients applied where appropriate to account for no additions taking place for annual treatments in 2001. Efforts were made to account for non-equivalence of N and P supply between FYM treatments and their corresponding inorganic tmts in mixed modelling statistics.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Mean outcome values per treatment per year were provided enabling trends in intervention effects to be seen. Overall ANOVAs provide p values showing significant effects of explanatory variables. Significant effects of individual treatments on response variables are described in the text.

Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++ And ☐+	Comments: Yes for lime treatments the experiment scores 1++ But due to less N and P being supplied in the inorganic treatments compared with their FYM comparators for first 7 years, score is reduced to 1+ for the nutrient addition aspect.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Yes, the findings can be extrapolated to MG3 and MG5 meadows with a similar management history. They should be interpreted carefully on MG3 sites with lower historic inputs.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Kirkham FW, Mountford JO & Wilkins RJ
	Year	1996
	Aim of study	To identify the extent and nature of botanical change at different levels of N,P and K applications - To ascertain how far cutting early at a high N rate (as for silage) influences botanical change compared with cutting for hay at the same N rate; - To determine whether the effects of fertilizers on species diversity can be mitigated by applying most or all of the annual N application after hay cutting - To compare results with the Large Scale Experiment to compare the findings with aftermath grazing versus cutting alone (two cut regime).
	Study design	1
	Quality score	++
	External validity	+ (Partially relevant due to study taking place on peat soils)
Population and setting	Source population	Species rich hay meadows of the NVC types MG5, MG8 and MG4.
	Eligible population	Species rich hay meadows

	Inclusion and exclusion criteria	
	Setting	Tadham Moor SSSI in the Brue Valley, Somerset Levels
Methods of allocation to intervention/control	Methods of allocation	Experiment replicate employed 3 blocks, in which 19 treatments were randomly allocated to plots within blocks
	Intervention description	1st Small scale experiment under cutting management only once after 1 July and again in autumn: • Fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Plus 100 or 200 kg N/ha with 0P and K replaced, 0 or 100 or 200 kg N/ha with 75kg P/ha and K replaced and 200kg N/ha, 75 kg P/ha and 200 kg K/ha. The effect of timing of fertilizer was investigated by applying N on up to four occasions each year - most treatments were split between spring and mid season . With the exception of treatment 8 which was cut first in mid May, all plots were cut first after the 1st July each year. 2nd Small scale experiment set up within N0 and N200 large scale plots in 1991 - 1992 to investigate influence of cutting date and previous fertiliser treatment – NB results not presented as outside scope of this review. Cutting dates were wither late May, early-mid July, early August or in early September, then aftermath grazed

	Control/comparison description	Yes - O NPK input control
	Sample sizes	24 x1m2 quadrats per plot (1986 – 1989)
	Baseline comparisons	1986 first year of experiment after set up. Second small scale experiment set up in 1991.
	Study sufficiently powered	Yes X 3 replication – sampling sufficient
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Botanical assessments were made in May and October each year. **% cover of species present** In first year 1986 – species cover abundance data were obtained from each plot using a vertical point quadrat to record all hits to ground level at 50 plots per point. These data were supplemented by visual scoring for the relative abundance of each species present within eight 20 x 20 cm quadrats per plot and the two datasets combined to give relative abundance values (% cover) for each species. Plots were assessed by visual scoring only on subsequent occasions using 0.5m2 quadrats. % cover of litter and bare ground Vegetation height Density of inflorescence of a number of species of conservation interest was recorded in late June each year.
	Secondary outcome measures	Species richness

		Simpson's index of diversity
	Follow-up periods	4 years
	Methods of analysis	Individual species abundance data, the number of species per plot, Simpson's Index and biomass data were analysed separately for treatment effects within treatment series by analysis of variance (ANOVA) Relative abundance data for each species were used to produce dominance –diversity curves for each treatment within the NPK series using the treatment means from the May assessment Two forms of ordination were used to relate community composition to N, P and K applications CCA and DCA.
Results		Treatments that included N applied at 25 kg ha-1 yr-1 with both P and K replaced at c 13 kg P ha-1 yr-1 and between 56 to 106 kg K ha-1 yr-1) significantly reduced ($p<0.05$) Simpson‟s index of botanical diversity compared to the control after just one yearWithin two years species diversity was significantly lower on plots receiving N100 with the high rate of P at 75 kg ha-1 yr-1 than on those receiving the same amount of N but with replacement P and K only ($p<0.01$). Ordination studies indicated that botanical change was in fact influenced to a greater extent by P than by N. Where P was applied without N changes in species richness and diversity were minimal even at the high application rate of 75 kg P ha-1 yr-1. Varying the proportions of the total N applied annually between spring and mid-summer (after hay cutting) had no significant effect on either species richness or species diversity of the vegetation. However, the authors suggest that this finding may be attributable to the overriding effect of replacing P and K, in both spring and mid-season. There was little difference in the pattern of dominance –diversity between the early cut versus normal cut treatments.
Notes	Limitations identified by author	Changes in the control plots are likely to have been attributable to the switch from grazing to cutting – absence of aftermath grazing within this experiment reduces its wider applicability. The experiment was not fully factorial - no treatment of P applied without K.

	Limitations identified by review team	Note this experiment was on a peat soil which are typically more deficient in plant available P compared to mineral soils.
	Evidence gaps and/pr recommendations for further research	Of those directly relevant to this project further research is needed to: Understand P availability and its effects on the recovery and maintenance of high floristic diversity Identify optimum conditions for the recruitment of seedlings of sensitive and/or rare species into these meadow communities, specifically by understanding the role of grazing.
	Sources of funding	MAFF, NCC and DOE

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Kirkham, F.W., Mountford, J.O., & Wilkins, R.J. 1996. The effects of nitrogen, potassium and phosphorus addition on the vegetation of a Somerset peat moor under cutting management. *Journal of Applied Ecology*, 33, 1013-1029.
Study Design Category	1
Assessed by & when	CE Pinches, 24th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Yes, very well described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Yes – meadows comprise MG4, MG5 and MG8 NVC communities
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Yes.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to	☐++	Randomised block design

management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?		
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Yes, comprehensively described in the paper.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Not reported so presumed to have been low
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	No interventions - other than those described. Minor deviations from intervention management, when hay cutting dates delayed in 1988 by a month due to bad weather.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Yes
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Yes

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable?	☐+	Subjective visual assessment of % cover .

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?		
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Yes

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Yes
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard.	☐++	No power calculation undertaken but degree of replication and design of experiment mean study has sufficient power.

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐++	Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Yes
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Yes statistical findings well reported.
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Yes
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Yes, but less so for Upland Hay Meadows as the Tadham study site overlies peat.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Kruk, M, Noordervliet, MAW & ter Keurs, WJ 1996. Hatching dates of waders and mowing dates in intensively exploited grassland areas in different years. *Biological Conservation* 77: 213-218.
	Year	1996
	Aim of study	To examine the relationship between mowing and hatching dates over an 8 year period as influenced by spring temperatures
	Study design	2
	Quality score	++
	External validity	+
Population and setting	Source population	655 grass fields
	Eligible population	Ade and Duivenvoorde, Netherlands
	Inclusion and exclusion criteria	Fields were cut for silage between April and June, then grazed later in year. Fields in management agreement were specifically excluded.

	Setting	Netherlands
Methods of allocation to intervention/control	Methods of allocation	Observational study
	Intervention description	NA
	Control/comparison description	NA
	Sample sizes	655 fields
	Baseline comparisons	-
	Study sufficiently powered	Yes
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Data on nest s of lapwing, black tailed godwit and redshank were collected weekly from the end of March until the beginning of June. Nests ultimately recorded as being – successful, preyed upon, destroyed by grassland operations or deserted since last visit. T-sums – from nearest weather station
	Secondary outcome measures	First egg date estimated Hatching date calculated
	Follow-up periods	8 years
	Methods of analysis	Spearman's rank correlation
Results		The date at which T sum of 180 C was reached varied between 30 Januray and 30 March. The warmer the spring the earlier the median mowing date in that year (p =0.04) Significant relationships were also found between no of days after 1st January when the T sum

		reached 180 C and the median hatching dates for black tailed godwits (p=0.031), redshanks (p =0.048) and a weaker correlation for lapwings (p =0.069). A both median hatching dates and median mowing dates are correlated with T sums a close relationship was found for all three species (p=<0.055 There are great differences in mowing and hatching dates between years which are determined by spring temperatures. Negative effects of early mowing on the breeding success of waders are as a consequence smaller than expected. Despite this in order to maintain the current populations levels mowing dates need to be delayed by 1 -2 weeks in order to ensure that the so called required recruitment must be met, a specific % which takes into account data in chick and adult survival. The difference between median mowing date in a particular year and the date for achieving the required recruitment showed that mowing date was too early in 5/8 years for lapwings, 4/8 for black tailed godwits and in 3/8 years for redshanks. These results suggest that safe mowing dates would have been 1-2 weeks later than current dates and that Tsum could be used to predict peak hatching for wader species to inform this in each year.
Notes	Limitations identified by author	None
	Limitations identified by review team	None
	Evidence gaps and/pr recommendations for further research	None
	Sources of funding	Leiden University, the Netherlands

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Kruk, M, Noordervliet, MAW & ter Keurs, WJ 1996. Hatching dates of waders and mowing dates in intensively exploited grassland areas in different years. *Biological Conservation* 77: 213-218.
Study Design Category	2
Assessed by & when	10th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	Comments: Management of grasslands studied relatively well described.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+/ NR	Comments: Yes
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Very large sample, selection based on cutting regime for silage.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	NA	Comments: Observational study
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	NA	Comments: Observational study NA
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: Yes, predation accounted for
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes in terms of typical management of silage fields.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Objective
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	Comments: Yes
3.4 Were outcomes relevant?	☐+	Comments: Yes, whilst this study focuses on intensively managed grassland impacts of field

Where surrogate outcome measures were used, did they measure what they set out to measure?		operations on clutch survival are relevant to the hay meadows questions as is relationship with T sum, peak hatching time and median mowing date.
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: Yes. 8 years.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: Very little information provided on analysis but these appear to all be non-parametric.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: No, but study tests simple hypothesis does T Sum effect hatching date and mowing date.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: yes, ok
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes, p values and r values givem ‘
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential	☐++	

confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Responses of different wader populations (i.e. in Northern Pennines) to T sun would need to be well characterised before management guidance could be set for birds on the basis of Tsum.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Lawes, J. B. , Gilbert, J. H. and Masters, M. T.
	Year	1882
	Aim of study	Original aim was to investigate methods of improving yields of hay and determine the effect of different fertilizer regimes on the yield of hay from permanent grassland.
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Lowland neutral grassland MG5
	Eligible population	As above
	Inclusion and exclusion criteria	
	Setting	2.85 ha of neutral grassland resembling NVC type MG5 Rothamstead, Hertfordshire
Methods of allocation to	Methods of allocation	The original experiment consisted of large plots 0.2 ha to which different fertilizers

intervention/control		are applied.
	Intervention description	Thirteen plots were originally established ranging in size from 0.05 and 0.2 ha. Initially each plot received either no nutrient addition (the control plots, Plot 3 and 12), Farm yard manure (Plot 2) at 14 tonnes per acre (35 tonnes ha$^{-1)}$ annually was included in the initial treatments but discontinued after eight years. Mineral fertiliser was applied as follows Nitrogen was applied annually in three amounts (48, 96 and 144 kg ha^{-1}) as ammonium sulphate and in two amounts as sodium nitrate (48 and 96 kg ha^{-1}) together with P supplied at 35 kg ha^{-1}and K supplied at 225 kg ha^{-1}. The annual application of 35 tonnes per ha FYM is (equivalent to 240 kg N ha^{-1}, 45 P kg ha^{-1} and 350 kg K ha^{-1} - as presented in Rothamstead report on Long Term Classical Experiments – http://www.rothamsted.ac.uk/resources/LongTermExperiments.pdf) thereby supplying higher rates of nutrients than the inorganic equivalents). The plots were cut in mid-June and made into hay. For the first 19 years the re-growth was grazed by sheep penned on individual plots but since 1875 a second harvest has been cut and removed immediately.
	Control/comparison description	Yes – Plots 3 and 12
	Sample sizes	Unreplicated
	Baseline comparisons	Uniformity of the site was assessed in the five years prior to 1856 (according to Rothamstead report on Long Term Classical Experiments - http://www.rothamsted.ac.uk/resources/LongTermExperiments.pdf)
	Study sufficiently powered	No.
Outcomes and methods of analysis (inc effect size, CIs	Primary outcome measures	Samples of hay were take for botanical analysis from all plots for the first time in 1862 – these samples were dried, separated into species and weighed to give and estimated of the absolute

for each outcome and significance)		and percentage composition of each species in the total cropped biomass. In the third year of the experiment, samples of the hay crop from seven of the most characteristically different plots were taken and separated into Gramineous herbage, Leguminous herbage and Miscellaneous herbage. From 1862 more complete botanical assessments were under taken - bulk samples of 10, 12.5 , 15 or 20 lbs were taken from the hay from the plots and a % dry weight of each species determined and division into Gramineous herbage, Leguminous herbage and Miscellaneous herbage as above.
	Secondary outcome measures	Hay yield
	Follow-up periods	Paper summarises findings of first 20 years of experiment
	Methods of analysis	None – results simply presented – pre dates era of modern statistics
Results		Poa trivialis and Bromus mollis became co-dominant with FYM application to plot 2 byut subsequently declined after application ceased mainly in favour of Agrostis capillaris, Festuca rubra, Helictotrichon pubesens and Holcus lanatus. Four years after cessation of FYM application the vegetation consisted of 85% by weight of grasses, 1.6% legumes and 14% others. These proportions were very similar to those on the plots receiving annually 48, 35 and 225 kg ha-1 NPK respectively and markedly different from the unfertilised plots (62% grasses, 8.1% legumes and 30% others averaged over two plots. In summary, both the FYM treatments and combinations of NPK with N applied at its lowest rate of 48 kg ha^{-1} annually, quickly caused significant change in the proportions of the grasses, legumes in the herbage. Nitrogen fertiliser suppressed legumes and other forbs and PK fertilisers without N encouraged legumes. Ammonium sulphate alone or with P K fertilisers

		eliminated the legumes, leaving a herbage with 90% or more grasses. It should be initial annual application rates of 35 tonnes/ha/year caused 'adverse effects' to the sward from smothering**.**
Notes	Limitations identified by author	-
	Limitations identified by review team	Due to age of experiment there was no randomization of treatments and replication is uneven. Park Grass plots were subject to aftermath grazing for the first 20 years, thereafter the aftermath was removed by cutting. The early FYM treatment included allows little comparison with the inorganic treatments due to non equivalence of the rates of nutrients applied (the annual FYM treatment supplied higher annual rates of all macro-nutrients N, P and K compared to the inorganic treatments).
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Lawes Trust

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Lawes, J. B. , Gilbert, J. H. and Masters, M. T. (1882) Agricultural, botanical and chemical results of experiments on the mixed herbage of permanent meadow, conducted for more than twenty years in succession on the same land. Part II The botanical results. *Philosophical Transactions of the Royal Society (A and B)*, 173, 1181-1413.
Study Design Category	2
Assessed by & when	CE Pinches, 24th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Not described in detail in this paper but described fully in other published literature and is MG5 grassland
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: MG5 grassland present on site is known to be representative of that type in lowland England.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	Comments: Means by which treatment plots were allocated is not described in this paper but is known to be non-random. The experiment is 150 years old so set up of treatment plots pre-dates modern concepts of good experimental design.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐+	Comments: Sampling comprised representative samples of differing weights taken from plot areas- % dry weight of each species determined
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Yes relatively - although impossible to know the amount of N,P and K actually supplied in teh FYM treatments.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Yes plots are >100m2
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐+	Comments: Park Grass plots were subject to aftermath grazing for the first 20 years on which this reference reports
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes species rich MG5 representative of wider UK species rich lowland meadow resource.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Broadly though treatments represent historical practice, for example use of ammonium N and fish meal.

Section 3: Outcomes

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Yes objective dry weight assessments of species composition.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐+	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes to the extent that they could be at the time
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐-	Comments: No since, after 8 years annual applications of FYM on plots 1 and 2 was ceased, although annual application of ammonia salts continued on plot 1. Inputs continued on all other plots – hence plots 2 was in effect recovering post 1863.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Yes study looks at first 20 years.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Plots were established in reputedly uniform botanical composition.

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: Replication is uneven.
4.3 Were the estimates of effect size given or calculable?	☐-	Comments: No
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐NR	Comments: Simple description of results presented – publication pre dates modern statistics. Crude differences similarities in treatment
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Yes these findings provide indication of the effect of different nutrient regimes on botanical species richness over a 20 year period.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Yes for MG5.

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Mercer, P., Reavey, C., Morgan, J.	**Source population:** MG10 (Ederney) and MG10/M23 (Teemore)	**Methods of allocation:** Three different concentrations of glyphosate in a hand-held weedwiper 0.5 m wide. Plants wiped first in one direction then the other. Ederney had area wiped in one/two years/not wiped. Teemore wiped in spring/early summer or late summer/autumn	**Primary outcome measures:** Changes in biomass/relative biomass of rushes, or changes in % cover of rushes	Weed-wiping in spring/early summer significantly reduced rush biomass and rush % cover the following year, but the effect was reduced after two years. The effect was much reduced in plots treated in the autumn. Different concentrations of glyphosate had no significant effect on rush biomass, % cover or % broad-leaved plants. Biomass of grass was also significantly reduced wherever biomass of rush was reduced.	**Limitations identified by author:**

Year: 2008 **Aim of study:** To investigate the effects of glyphosate application on rush growth, incliding timing, treatment **Study design:** Randomised block **Quality Score:** + **External validity:** +	**Setting:** Co. Fermanagh, Ireland	**Intervention description:** 4 replicates at each site, arranged in randomised blocks, plot size 5 m x 7 m (Ederney) or 5 m x 10 m (Teemore) **Control / comparison description:** A control treatment is mentioned for the Ederney site, but not sure if there was one control block or four replicates. Only one control treatment at Teemore **Baseline comparisons:** No species data taken before the start of weed-wiping. Biomass measurements made in autumn of 1st, 2nd and 3rd year (Ederney) and 2nd, 3rd and 4th years (Teemore). Point quadrat measurements in autumn of 2nd year (Ederney) and 3rd/4th years (Teemore)	**Secondary outcome measures:** **Follow-up periods:** Different for the two **Methods of analysis:** Analysis of Variance		**Evidence gaps and recommendations for further research:** A comparison between weed **Sources of funding:** Authors employed by the Agri Food and Biosciences Institute/Environment and Heritage Service

Overall score: 1+		**Study sufficiently powered:** No power analysis - very likely not sufficiently powered			

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	b - management methods to control rushes
Study Citation	Mercer, P., Reavey, C., &Morgan, J. (2008). Control of *Juncus* spp. in grassland similar to Environmentally Sensitive Areas in Northern Ireland, using a weed-wiper
Study Design Category	Randomised block design
Assessed by & when	Kate Fagan 13/11/12

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Article gives approximate NVC types, also proportions of rushes, and describes roughly where they are situated

1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	The article states that they are typical of the area
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	No method of selection described, but randomised blocks used

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Randomised block design
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Treatments well described and appropriate
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias?	☐++	Consistency good - all dates of exposures given

Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)		
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	No contamination reported
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	o-	At Teemore, some plots were treated only one year, others for two consecutive years and other for three, and these differences didn't inform the results
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Yes, and approximate NVC types were given
2.7 Did the intervention(s) or	☐+	The method didn't - they used a hand-held weed-wiper. But weed-wiping is

control comparison(s) reflect the usual UK practice(s)?		a realistic management practice

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Objective outcome measurements
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Yes

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments:

3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	The Teemore treatments lasted for different amounts of time, so the post-treatment time intervals differed. The differences were ignored in analysis.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	A longer time interval would have been useful, to establish whether two-years of teatment had a longer-lasting effect than one year of treatment

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐-	No baseline data

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	No power calculation, replication low
4.3 Were the estimates of effect size given or calculable?	☐++	All stats given

4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	**Comments**:
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐++	All stats given
[redacted]		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐+	Experimental design good

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++	Yes

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Year: **Aim of study:**	**Source population:** **Setting:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** Limitations identified by review team: Failure to control for effect of baseline vegetation composition of plots in detailed comparison of species composition and species attributes between treatments in 1991. Baseline vegetation shoudl have been treated as a covariate.

Study design: **Quality Score:** **External validity:**		**Control / comparison description:** **Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Follow-up periods:** **Methods of analysis:**		**Evidence gaps and recommendations for further research:** **Sources of funding:**
Overall score:					

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?**	☐++	Comments:

Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?		

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likelv/not likelv?	☐++	Comments:

2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments:
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments:
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Comments:

2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective.	☐++	Comments:

How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?		
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments:
3.3 Were all important outcomes assessed?	☐++	Comments:

Were all important positive and negative effects assessed by the variables/measurements used?		
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:

3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments:

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	**Comments**:

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	**Comments**:
4.3 Were the estimates of effect size given or calculable?	☐+	**Comments**:

4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	**Comments**:
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐+	**Comments**:

[redacted]

5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments:
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments:

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R. **Year:**	**Source population:** **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** **Limitations identified by review team:**

Aim of study:					**Evidence gaps and recommendations for further research:**
Study design:	**Setting:**	**Control / comparison description:**	**Follow-up periods:**		

Quality Score: **External validity:**		**Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Methods of analysis:**		**Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	2
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?**		Comments:

e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled?		

Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	
2.5 Is the setting applicable to the UK?	☐+	

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	Comments:
3.2 Were all outcome measurements complete?		

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐-	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	

3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	

4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	

5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Year: Aim of study: Study design: Quality Score: External validity: Overall score:	Source population: Setting:	Methods of allocation: Intervention description: Control / comparison description: Baseline comparisons: Study sufficiently powered:	Primary outcome measures: Secondary outcome measures: Follow-up periods: Methods of analysis:		Limitations identified by author: Limitations identified by review team: Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review
Name of Review Sub-topic (if any):

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments:
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions?	☐ Clear	Comments:

-is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?		
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:

2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question?		Comments:

-Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	

3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Not Sure / inadequately reported	Comments:

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Clearly described	Comments:
4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments:
4.3 Were the methods reliable? For example: -was data collected by more than one method?	☐ Reliable	Comments:

-is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?		

5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments:
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:

5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	☐ Not sure / not reported	Comments:
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not Sure	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Partially relevant	Comments:

5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	Comments:
6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ Appropriately	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ +	Comments:

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Mountford, J.O., Lakhani, K & Kirkham
	Year	1993
	Aim of study	To examine the effects of a wide range of fertilizer treatments on species diversity, agricultural production and losses of soil N in these meadows
	Study design	1
	Quality score	++
	External validity	+ (Partially relevant due to study taking place on peat soils)
Population and setting	Source population	Species rich hay meadows of the NVC types MG5, MG8 and MG4.
	Eligible population	Species rich hay meadows
	Inclusion and exclusion	

	criteria	
	Setting	Tadham Moor SSSI in the Brue Valley, Somerset Levels
Methods of allocation to intervention/control	Methods of allocation	Experiment employed 3 blocks, in which 5 treatments were randomly allocated to plots within blocks
	Intervention description	• Five fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Phosphorous (as Triple Phosphate) and Potassium (muriate of Potash) were applied in amounts to replace that removed in the hay crop on all plots except controls - calculated from yield and chemical analysis of hay swath samples. • Annual applications of N were split between two equal dressings, the first as soon as ground conditions allowed after mid April and the second after the removal of the hay crop. P and K were applied in mid season each year on the day following the second N application. • Treatment plots were cut for hay after July 1st and the aftermath grazed by beef cattle – a compressed sward height of5.5-6.5cm was maintained during grazing period.
	Control/comparison description	O NPK input control
	Sample sizes	24 x1m$_2$ quadrats per plot (1986 – 1989)
	Baseline comparisons	1986 first year of experiment after set up.
	Study sufficiently powered	Yes X 3 replication – sampling sufficient

Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	**Botanical** % cover of species present % cover of litter and bare ground Vegetation height Density of inflorescence of a number of species of conservation interest was recorded in late June each year.
	Secondary outcome measures	Mean species number per quadrat in each plot Species richness Flowering species richness Simpson's index of diversity
	Follow-up periods	5 years 1986 – 1990.
	Methods of analysis	ANOVA of each variable in each year to test the null hypothesis of equality of the experimental treatments. If the null Ho rejected then each of the 4 nitrogen application treatments was compared with the control treatment, using student's t test. Significance of linear effects of nitrogen levels was also examined for every variable.
Results		**Only the effects on botanical composition are reported**

		Headline findings after 5 years – An annual application of 25 kg N/ha/yr stimulate the spread of agriculturally productive grasses within 2 years and 50 kg n/ha/yr significantly reduced species richness in three years. Significant reduction in species number occurred within 2 years under inputs of 100 or 200kg N per ha, 3 years with inputs of 50kg N per ha Four grasses (*Holcus lanatus, Lolium perenne, Phleum pratense* and *Poa trivialis*) showed a positive linear trend with nitrogen. The trend in *H. lanatus* and *L.perenne* became increasingly significant with time, these two grasses dominating the plots receiving high rates of N. However in 1987 a significant effect of the N25kg treatment was seen for *Holcus lanatus* and by 1988 for *Lolium perenne*. After 5 years Anthoxanthum odoratum (a key grass of UHMs) showed no significant trends in response to N. The majority on non-grass species showed a negative linear trend with nitrogen, which often became more significant in later years. All three rush species recorded showed this pattern as did most sedges and many low growing forbs. Form 1987 some forbs and mosses had significantly lower treatment means under the N25 and N50 treatments compared to the controls. There were significant reductions to the number of species in flower in the 50, 100 and 200kg N plots. Vegetation height showed a positive linear trend with nitrogen applied that became most significant in 1989 and then less so in 1990 following winter floods and a very dry spring.
Notes	Limitations identified by author	Site flooded in March 1990. Small plot experiment was not fully factorial - no treatment of P applied without K.
	Limitations identified by review team	Note this experiment was on a peat soil which are typically more deficient in plant available P compared to mineral soils.
	Evidence gaps and/pr recommendations for	Of those directly relevant to this project further research is needed to:

	further research	Understand P availability and its effects on the recovery and maintenance of high floristic diversity Identify optimum conditions for the recruitment of seedlings of sensitive and/or rare species into these meadow communities, specifically by understanding the role of grazing.
	Sources of funding	MAFF, NCC and DOE

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Mountford, J.O., Lakhani, K & Kirkham
	Year	1993
	Aim of study	To examine the effects of a wide range of fertilizer treatments on species diversity, agricultural production and losses of soil N in these meadows
	Study design	1
	Quality score	++
	External validity	+ (Partially relevant due to study taking place on peat soils)
Population and setting	Source population	Species rich hay meadows of the NVC types MG5, MG8 and MG4.
	Eligible population	Species rich hay meadows
	Inclusion and exclusion	

	criteria	
	Setting	Tadham Moor SSSI in the Brue Valley, Somerset Levels
Methods of allocation to intervention/control	Methods of allocation	Experiment employed 3 blocks, in which 5 treatments were randomly allocated to plots within blocks
	Intervention description	• Five fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Phosphorous (as Triple Phosphate) and Potassium (muriate of Potash) were applied in amounts to replace that removed in the hay crop on all plots except controls - calculated from yield and chemical analysis of hay swath samples. • Annual applications of N were split between two equal dressings, the first as soon as ground conditions allowed after mid April and the second after the removal of the hay crop. P and K were applied in mid season each year on the day following the second N application. • Treatment plots were cut for hay after July 1st and the aftermath grazed by beef cattle – a compressed sward height of5.5-6.5cm was maintained during grazing period.
	Control/comparison description	Large scale experiment O NPK input control
	Sample sizes	24 x1m$_2$ quadrats per plot (1986 – 1989)
	Baseline comparisons	1986 first year of experiment after set up.
	Study sufficiently powered	Yes X 3 replication – sampling sufficient

Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	**Botanical** % cover of species present % cover of litter and bare ground Vegetation height Density of inflorescence of a number of species of conservation interest was recorded in late June each year.
	Secondary outcome measures	Mean species number per quadrat in each plot Species richness Flowering species richness Simpson's index of diversity
	Follow-up periods	5 years 1986 – 1990.
	Methods of analysis	ANOVA of each variable in each year to test the null hypothesis of equality of the experimental treatments. If the null Ho rejected then each of the 4 nitrogen application treatments was compared with the control treatment, using student's t test. Significance of linear effects of nitrogen levels was also examined.
Results		**Effects on botanical composition reported only** **Large scale experiment**

		Species richness of the hay meadows was significantly lower than the control in the lowest fertilizer input of 25kg N per ha per annum within 6 years. Significant reduction in species number occurred within 2 years under inputs of 100 or 200kg N per ha, 3 years with inputs of 50kg N per ha There were significant reductions to the number of species in flower in the 50, 100 and 200kg N plots. A taller grass dominated sward was created in plots that received 50kg or more of N per ha, Lolium perenne was the dominant species on all fertilized plots. **Species changes** Of the 157 species recorded in the study area between 1986 and 1993, the abundance of 50 as significantly affected by fertilizer treatment in at least one year. Of these 13 species showed a significant increase – Agrostis stolonifera, Bromus hordaeceus, Bromus racemosus, cerastium fontanum, Cirsium arvense, Holcus lanatus, Lolium perenne, Phleum pratense, Poa trivials, Rumex acetosa, Rumex crispus, Stellaria media and Taraxacum agg. R. Acetosa is known to be stimulated by P, ammonium and organic fertilizers but discouraged by nitrates, the increase may be due to the application of replacement P and K, this was supported by findings from the small plots. Some low growing forbs and bryophytes disappeared locally in high N treatment plots. A large number of forb species showed a significant reduction in abundance on plots receiving fertilizer. 44 species showed a significant decrease in abundance in response to fertilizer input in at least one year, six of these were grasses, 6 were sedges and the rest were lower growing dicotyledonous species and mosses. These lower growing species were effectively being competitively excluded from the fertilized plots, by the addition of the fertilisers stimulating earlier growth and shortening the period before which lightlyavailability became severely limiting. Vegetation height showed a positive linear trend with nitrogen applied. The number of flowering plants of species indicative of old wet meadows declined in response to fertilizer input. Meadow thistle, Cirsium dissectum, Ragged robin, Lychnis flos cuculi, Cuckoo flower Cardamine pratensis, Lotus pedunculatus and Meadowsweet Filipendula ulmaria almost

		completely disappeared in plots receiving high N inputs. Prior to 1986 these were all abundant but, after 7 years of fertilizer applications they were only common on the control plots receiving no inputs.
Notes	Limitations identified by author	Site flooded in March 1990. Small plot experiment was not fully factorial - no treatment of P applied without K.
	Limitations identified by review team	Note this experiment was on a peat soil which are typically more deficient in plant available P compared to mineral soils.
	Evidence gaps and/pr recommendations for further research	Of those directly relevant to this project further research is needed to: Understand P availability and its effects on the recovery and maintenance of high floristic diversity Identify optimum conditions for the recruitment of seedlings of sensitive and/or rare species into these meadow communities, specifically by understanding the role of grazing.
	Sources of funding	MAFF, NCC and DOE

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Mountford, J.D., Lakhani, K.H., & Kirkham, F.W. 1993. Experimental assessment of the effects of nitrogen addition under hay-cutting and aftermath grazing on the vegetation of meadows on a Somerset peat moor. *Journal of Applied Ecology*, 30, 321-332.
Study Design Category	1
Assessed by & when	20th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Yes, very well described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Yes
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Yes.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to	☐++	Randomised block design

management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?		
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Yes, comprehensively described in the paper.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Yes
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	None other than those described.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Yes
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Yes

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable?	☐+	Subjective visual assessment of % cover .

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?		
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Yes

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Yes
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard.	☐++	Yes

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐++	Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Yes
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Yes
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Yes
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Yes, but less so for Upland Hay Meadows as the Tadham study site overlies peat.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Maria Pacha and Sandrine Petit **Year:** 2008	**Source population:** Uplands hay meadows in the Yorkshire Dales National Park **Eligible population inclusion & exclusion criteria:**	**Methods of allocation:** A sub-sample of 47 fields were surveyed through 1 x 1 m quadrats for plant presence and abundance. The remainder were surveyed across transects for presence only. The selection of the 47 wasn't described **Intervention description:** Not applicable for this study	**Primary outcome measures:** Plant species presence/abundance **Secondary outcome measures:** Habitat quality index	Presence of G. sylvaticum declined by 40% between the two survey periods. The variables best explaining the decline in G. sylvaticum were declining habitat quality and site isolation. Meadow quality declined	**Limitations identified by author:** **Limitations identified by review team:**

Aim of study: To investigate the changes in the vegetation of upland hay meadows over the last two decades, and how these changes were related to management practices and isolation, particularly concentrating on Geranium sylvaticum	Meadows that had been survey during the 1980s were selected depending on the presence of Geranium sylvaticum		devised using presence/absence information from a sub-sample of fields (strongly based on *Geranium sylvaticum* information)	significantly between the two survey periods with declines in species richness ($p<0.01$) and a 40% loss of sites supporting wood crane's bill G.sylvaticum. Species richness was found to be negatively correlated with high grazing intensity ($p<0.01$) and inorganic fertiliser application ($p<0.01$). Meadow quality, as described by a derived habitat	Methods weren't always well described. Survey techniques for the two different periods aren't clear, nor is site selection. The way in which habitat management categories fitted into results isn;t always clear (for example the apparent 3 fertilisation categories mentioned in the methods become just 'fertiliser
Study design: Re-survey of all sites identified meeting the criteria within a specified area, with corresponding survey of management practices/site isolation **Quality Score: ++** **External validity:** ++	**Setting:** Yorkshire Dales National Park	**Control / comparison description:** Not applicable for this study **Sample sizes:** Total of 119 fields **Baseline comparisons:** Surveys surveyed in the 1980s **Study sufficiently powered:** There is no power calculation	**Follow-up periods:** Initial surveys in the 1980s, revisited in 2003 **Methods of analysis:** A combination of Chi-squared association analysis and constant and characteristic species from the relevant NVC table		**Evidence gaps and recommendations for further research:** **Sources of funding:** Lancater University and Fundacion Jose Estensoso (Repsol-YPF), Argentina

		but the number of sites is large. There is no information on the number of sites in each of the management practice categories, however	were used to produce the habitat quality index. Spearman's correlation was used to assess the effect of management practices on the quality index. Stepwise logistic regression General Linear Modelling was used to predict the presence of *Geranium sylvaticum* .		

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	c (a) - methods that best maintain floristic diversity
Study Citation	Pacha, M. & Petit, S. (2008). The effect of landscape structure and habitat quality on the occurrence of Geranium sylvaticum in fragmented hay meadows. Agriculture, Ecosystems and Environment 123, 81-87
Study Design Category	2
Assessed by & when	Kate Fagan 25/11/12

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Comments:
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	It appears that all of the eligible sites were surveyed
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	The selection of the 47 sub-sampled fields used for the development of the habitat quality index wasn't described.

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Well discussed
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled?		No confounding factors mentioned, and the amount of variation explained suggests that confounding factors weren't important

Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++	
2.5 Is the setting applicable to the UK?	☐++	Yorkshire Dales National Park

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Insufficient information given on survey methodology, but this is unlikely to have an effect on the results
3.2 Were all outcome measurements complete?	☐++	

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	The habitat quality assessment appeared to work well
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	

3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐++	Unusually long time for such a study makes this study particularly valuable

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Presumably yes, but there is no information on how many fields fell into the different categories
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	
4.3 Were the analytical methods appropriate?		

Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Test statistics given
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++	
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++	

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Various – Details of references from the Park Grass Experiment that were evaluated for this review are set out below this table.
	Year	References span 1859- 2005
	Aim of study	**Original purpose** To investigate ways of improving the yield of hay by the application of inorganic fertilisers and organic manure. The experiment has subsequently provided the opportunity: to examine the continuing effects of the original treatments on species diversity and on soil function to tests of effects of different liming regimes **Specific aims of key references:** **Dodd et al. 1994** - explored the temporal aspect of community composition between 1856 and 1992, how quickly plots fertilised in a variety of ways lost or changed their original classification.
	Study design	2 (unreplicated)
	Quality score	

	External validity	
Population and setting	Source population	Lowland neutral grassland
	Eligible population	As above
	Inclusion and exclusion criteria	
	Setting	2.85 ha of neutral grassland resembling NVC type MG5 Rothamstead, Hertfordshire
Methods of allocation to intervention/control	Methods of allocation	The original experiment consisted of large plots to which different fertilizers are applied. In 1903 most plots were halved and the effects of regular liming tested. This was modified in 1965 with the division of most plots into four sub-plots, three of which are limed to maintain pHs of 5, 6 and 7. The fourth sub-plot receives no lime.
	Intervention description	For full details of treatments and experimental layout please refer to Silvertown et al. 2006, p.g 4 http://www.open.ac.uk/science/biosci/personalpages/j.silvertown/pdfs/Silvertown_et_al_2006.pdf NPK Various combinations of inorganic fertilisers (P, K, Mg, Na, nitrate-N, ammonium-N and Si) have been tested since the start; Lime Since 1903 the effect of lime has been tested. Lime applied every 3rd year Ground chalk applied as necessary to maintain the soil at pH7,6,5 on sub plots a,b,c respectively with sub plot d representing the nil input control. FYM Between 1856 -1863 FYM was applied annually to plot 2 in Nov/Dec at a rate of 35t/ha-1 but was

		discontinued after eight years because, when applied annually to the surface in large amounts, it had adverse effects on the sward. In 1905 FYM treatments were introduced on three plots, it was the applied every four years at a rate of 35 t per ha, supplying 240kg N, 45 kg P and 350kg K. The plots are cut in mid-June and made into hay. For the first 19 years the re-growth was grazed by sheep penned on individual plots but since 1875 a second harvest has been cut and removed immediately.
	Control/comparison description	Yes untreated plot 3
	Sample sizes	Unreplicated apart from 2 control plots, one of which may have been levelled initially using soil from elsewhere so control is also only 1x replication.
	Baseline comparisons	Yes, 1856, uniformity of the sward was assessed in the 5 years prior to treatments being applied.
	Study sufficiently powered	No.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Botanical composition of the plots has been recorded at irregular intervals with some substantial gaps/ Samples taken between 1862 and 1976 were from 36m2 cut areas - % dry weight of each species determined From 1991 to 2000, 6 randomly located quadrats measuring 50 cm x 25 cm were located within each plot in early June, vegetation was harvested and dry weight per species determined. When the 6 quadrats were aggregated this gave a measure of species richness at 0.75m2 for each plot. Whole species density at each plot was visited monthly from April to November each year and a composite list of species was compiled.

	Secondary outcome measures	Dodd et al. 1994 – Used MATCH to match plant communities found in treatments plots to NVC communities/sub communities.
	Follow-up periods	150 + years
	Methods of analysis	Dodd et al. 1994 chose 4 periods to analyse, for which there were a variable no of samples. 1. 1867 – 1877 2. 1900 – 1929 3. 1930 – 1949 4. 1973 – 1992 Data were assigned to NVC types using MATCH and the key to MG grasslands from Rodwell Crawley et al. 1994 applied a maximal model (including interaction terms and quadratic terms for continuous explanatory variables) was fitted first then the model simplified involved deletion of variables and reduction of factor levels. Explanatory variables are: experimental treatments: categorical variables with two levels in the case of P and K (applied or not); 3 levels for the type of N (none, ammonium sulphate, or sodium nitrate); 4 levels for liming; two levels for the transients; two levels for the organics (organics applied or not) and one continuous explanatory variable (application rate of N) with two covariates; total first cut biomass and soil pH. HH
Results		Botanical composition according to NVC type Plots receiving nitrogen free treatments moved from MG5b to MG1 e in 50 to 80 years. Plots which received nitrogen moved towards MG1 then to MG7d. Initial impacts of treatments within first 8 years. Fertilisers quickly changed the proportions of the grasses, legumes and weeds in the herbage. Nitrogen fertiliser suppressed legumes and weeds(other herbs) and PK fertilisers without N encouraged legumes. Lawes & Gilbert 1859 reported 2 years from the start of the experiment, PK fertilisers increased the legumes from 5 to 20% (dry weight mass) and all non-legume forbs were

		rare. Ammonium sulphate alone or with P K fertilisers eliminated the legumes and most of the weeds, leaving a herbage with 90% or more grasses. These large initial differences in the proportion of the 3 main groups of plants have persisted throughout the duration of the experiment, but changes have occurred in the composition of the groups themselves (Thurston, Williams & Johnston,1976). Following initial applications of FYM at 35 t/ha per year Poa trivialis and Bromus mollis became dominant. Four years after the cessation of FYM application the vegetation consisted by dry weight of 85% grasses, 1.6% legumes and 14% others. These proportions were very similar to those on the plot receiving N, P and K annually at 48,35 and 225kg/ha-1 respectively and markedly different from the vegetation on unfertilised plots (62% grasses, 8.1% legumes and 30% others. Total species no differed too, at 47,39 and 34 species per plot for unfertilised, FYM and NPK treatments respectively (note that the NPK was twice the area of the other two and this difference has not been controlled for). Subsequent impacts Species richness was greatest on plots that had no experimental inputs >40 and lowest in plots were the soil was strongly acidified by the long term input of ammonium sulphate supplying 144 N kg per ha. Species richness declines from the control plots, through plots receiving P alone, sodium nitrate or ammonium sulphate on their own, N and K together (-P), FYM and P together with K. The largest reduction in species richness are associated with adding N and P together and maximum depression of species richness occurs when N is applied as ammonium sulphate. Only N ($p<0.00001$) and P ($P<0.00001$) had significant main effects on species richness. There was no significant interaction between N and P application ($p=0.14$) the effect of adding N and P together was additive and was responsible for the greatest reduction is species richness attributable to nutrients. There was a roughly linear decline in mean species richness with N application rate for both types of N.

		Modern species numbers vary from 3 to 44 per 200 m2 among the plots (Crawley *et al* . 2005). According to the multivariate model of species density variation of Crawley *et al*. (2005), 50 kg N ha−1 year−1 added as fertilizer reduces species number by about 6.5 species, ammonium N loses 3 more species than would the same rate of N as sodium nitrate (because of the effect on soil pH), using organic manures rather than mineral fertilisers adds two species on average. Crawley 2004 showed that the addition of phosphorous reduced species richness, and application of potassium along with phosphorous reduced species richness further, but the biggest negative effects were when N and P were applied together. Liming There was no response to relationship between lime treatment and species richness except in plots receiving nitrogen in the form of ammonium sulphate, where species richness increased sharply with increasing pH (Crawley et , 2005). Another critical determinant of the species composition of the plot is the N:P ratio. A loss of species following the cessation of aftermath grazing was evident on all plots, including the control. For 15 or the first 21 years, until 1877, plots were grazed after hay cutting and the number of species recorded on control plots remained remarkably constant at about 50 (Lawes et al. 1882). The plots were not grazed after 1877 and the number of species declined progressively thereafter to an average of 37 (Williams 1978). This change was accompanied by a decrease in the fraction of grasses and a tendedncy for L.hispidis, P. Lanceolata and Sanguisroba officinalis to dominated. These findings are supportive of the key role of grazing in maintaining maximum diversity. Lime had only a small effect on the botanical composition of the unfertilised control (Thurston, Williams and Johnston 1976)
Notes	Limitations identified by author	Dodds et al. 1994 Only one sample per treatment per year available for analysis for comparison with the tables in the NVC, where ideally constancy of species between samples is required.

		Crawley et al. 2004 No randomization, replication is uneven, treatment combinations are missing and lime treatments are confounded with spatial location.
	Limitations identified by review team	Park Grass plots were subject to aftermath grazing for the first 20 years, thereafter the aftermath was removed by cutting. Botanical analysis of the 3 post 1905 FYM plots difficult to describe because two of them also receive fertilisers or fish guano. Only plot 19 is FYM only and a valid comparator.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NERC, BBSRC and Lawes trsut

References

Crawley, M.J.; Johnston, A.E.; Silvertown, J.; Dodd, M.; de Mazancourt, C.; Heard, M.S.; Henman, D.F. and Edwards, G.R. (2005). Determinants of species richness in the Park Grass experiment. American Naturalist, 165(2), pp. 179–192.

Dodd, M.E, Silvertown, J., McConway, K., Potts, J. & Crawley M (1994) Application of the British NVC to the communities of the Park Grass experiment through time. Folia Geobotanica et Phytotaxonomica , Praha 29; 321-224.

Warren, R.G. & Johnston, A.E. (1963) Rothamstead Experimental Station. Report for 1963. Lawes Agricultural Trust. Harpenden, Herts.

Thurston, J.M.; Williams, E.D.; Johnston, A.E. (1976) Modern developments in an experiment on permanent grassland started in 1856: Effects of fertilisers and lime on botanical composition and crop and soil analyses. Annales Agronomiques, 27 (5-6), p 1043-1082.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: Upland Evidence Review

Name of Review Sub-topic (if any):Hay Meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Crawley, M.J.; Johnston, A.E.; Silvertown, J.; Dodd, M.; de Mazancourt, C.; Heard, M.S.; Henman, D.F. and Edwards, G.R. (2005). Determinants of species richness in the Park Grass experiment. American Naturalist, 165(2), pp. 179–192.
Study Design Category	2
Assessed by & when	CE Pinches, 25h November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Yes, post hoc analysis of initial vegetation by Dodds et al. 1994 confirms that the meadow conforms to MG5.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes, presumed to be representative of neutral grasslands at the time 1856.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Yes, care was taken to check uniformity of the sward over the experimental site prior to setting up plots.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐-	Comments: Non random, plot size quite large .
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Yes, generally well described although some inconsistency in description of frequency of initial FYM treatment.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: Yes.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: Subsequent FYM treatments introduced in 1905 were sited on plots previously subject to inputs of NPK since 1972 potentially confounding findings.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes, hay cut and aftermath grazing initially then hay cut only after first 20 years.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Yes typical of species rich MG5 grassland.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Yes, at that time.

Section 3: Outcomes

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Botanical assessment by dry weight not recorded at same times in each plot particularly in initial phases of experiment which makes direct comparison difficult. Subsequent post 1999 botanical assessments standardised and more frequent.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐-	Comments: No as above.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments:
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	NA	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Broadly, though note newer FYM treatments started in 1905 not 1856!
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Yes for initial FYM treatments and inorganics applied in 1856, potentially differences exist for FYM treatments established on NPK plots in 1905. However initial findings suggesting comparable effects of first FYM treatments with inorganics may make this less of an issue.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐-	Comments: No

A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐-	Comments: yes for some studies
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	Comments: Yes in modern day studies evaluated Dodds et al. 1994 and Crawley et al. 2004
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes as above.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Yes, one of the most important long term experiments, unfortunately only one post 1905 FYM treatment is pertinent to this review.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++	Comments: Yes

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Anon - RSPB **Year:** 2007 for Information and Advice note **Aim of study:** Formal consensus rather than a study. Aim to give advice for management of rush-infested grassland for conservation management	**Source population:** The documents consider all grasslands with rush infestation of more than a third of the cover **Setting:** UK	**Methods of allocation:** NA **Intervention description:** No intervention	**Primary outcome measures:** NA **Secondary outcome measures:** NA	The guidelines advise a summer cut, after the last wader chicks have fledged (exact timing is dependent on species present) which should be as close to the ground as possible without causing bare soil which allows rush seeds in the seed bank the chance to establish. It is suggested that this will be more effective if followed after 4-8 weeks by another cut. Use of grazing as a management tool to control rushes is suggested, with grazing following a single cut reported as being sufficient in certain instances. Cattle are reported to be better than sheep at	**Limitations identified by author:** None **Limitations identified by review team:** No primary sources provided **Evidence gaps and recommendations for further research:**

Study design: Formal consensus **Quality Score: -** **External validity: -**		**Control / comparison description:** NA **Baseline comparisons:** NA **Study sufficiently powered:** NA	**Follow-up periods:** NA **Methods of analysis:** No analysis	suppressing rushes. Creeping rushes, namely (articulated rush and sharp flowered rush) are reported as being more readily grazed than tussock rushes (hard, soft and compact). The RSPB guidelines also mention the use of herbicide, specifically MCPA and glyphosate, as a possible rush control mechanism, using a weed-wiper, but warn of the likelihood that it will kill non-target vegetation unless there is a significant height difference between this and the rushes. Care is also advised in avoiding	**Sources of funding:** RSPB

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any):

Review Question	b - measures for controlling rushes (a - with consideration of how this affects breeding waders)
Study Citation	Anon (2007). Rush

	Management. http://www.rspb.org.uk/Images/rush_england_tcm9-207540.pdf. RSPB.
Study Design Category	4 - formal consensus
Assessed by & when	Kate Fagan 30-11-12

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	A quantitative approach, or at least reference to primary literature, would have better addressed the question at hand but would not have suited the purpose of informing land managers succinctly
1.2 **Is the study clear in what it seeks to do?** For example:	☐ Unclear	Aims not mentioned

- is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?		
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ NA	No reseach design - formal consensus of subject knowledge

2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ NA	No reseach design - formal consensus of subject knowledge

3.1 How well was the data collection carried out? For example:		No data collection

-Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ NA	

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐NA	No intervention, no researcher
4.2 Is the context clearly described?	☐NA	No observations

For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?		
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ NA	No data collection or any kind of investigation

5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit?		No data analysis

-how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ NA	
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ NA	No data
5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved?		No analysis

-were negative / discrepant results addressed?	☐ NA	
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not sure	No references given, no author so not sure of credentials, but RSPB so likely to be a reliable advice
5.5 Are the findings relevant to the aims of the study?	☐ Relevant	Comments:
5.6 Conclusions For example:		The whole publication is made up of conclusions

-how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ NA	but without any justification

6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ NA	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ -	Impossible to be confident of the reliability of the guidance

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	**What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?**

Study details	Authors	Shrubb, M.
	Year	1990
	Aim of study	To determine the impact of agricultural change on the nesting of Lapwings in England and Wales between 1962 and 1985 by analysis of BTO nest record cards. N.B for the purposes of this review only the aspects of this study that focus on evidence for the impact of relevant grassland management interventions on lapwing nesting success.
	Study design	2
	Quality score	++
	External validity	++
Population and setting	Source population	Sample of lapwing population across England and Wales, as recorded by Common Bird Census recording on grass, fallow (tilth) and arable.
	Eligible population	As above
	Inclusion and exclusion criteria	

	Setting	England and Wales
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	The cards analysed here were mainly collected between 1962 and 1985, allowing comparison with Lapwing populations monitored by the concurrent CBC. Some additional comparative analyses were made of cards collected from 1940-1961. Nesting habitats were classified under 10 agricultural categories: grassland was divided into upland rough grazings, upland improved grass, lowland rough grazing and lowland improved grass.
	Sample sizes	1093 nests were observed in Upland rough grass and 847 nests were observed in improved grass during the entire 1940 – 1961 period.
	Baseline comparisons	1940
	Study sufficiently powered	Correlative study .
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Nesting incidence and success - Only cards with a minimum of 2 visits spaced at least 4 days apart have been used to investigate clutch and brood size. Nests were accepted as successful when the brood was seen or the behaviour of the adults indicated the presence of a brood or the observer recorded evidence of a successsful hatch, usually the presence of hatched shells. Nest failure was accepted when the timing of visits showed that an empty nest could not have hatched or the observer recorded evidence of robbery/predation, destruction or desertion. The percentage of successful nests was calculated from all cards for which a definite result was known.
	Secondary outcome measures	-
	Follow-up periods	Study looks at records recorded from 1962 to 1985.

	Methods of analysis	The principal causes of nest loss by Lapwings nesting in different farmland habitats including grassland were examined between 1962 to 1985. The figures include all nests for which a result was known irrespective of whether they yielded information used to calculate clutch and brood size or not Simple statistics have been employed and presented – the details of the analysis are not described.
Results		Overall nesting performance in upland grass is now very poor because of greatly increased stocking rates. Stocking densities in England and Wales increased by a total of 37% over the study period, providing a consistent background to rising nest losses to trampling Not only are more nests lost to trampling and grassland cultivations, but increased numbers of cattle also cause more desertions. The study found that the percentage of lapwing nests in grassland lost to trampling in any year was significantly correlated with the overall densities of both sheep (rs =0.58, $P < 0.01$)and cattle (years 24, rs = 0.63, $P < 0.01$) on English and Welsh grasslands. The rate of nest desertion in grass also correlated positively with cattle densities (rs = 0.37, $P < 0.05$), as did losses to farmwork in grass (rs = 0.35, $P < 0.05$), but these factors did not relate to sheep numbers. Cattle farming in the uplands comprises proportionately more beef and stock rearing and less dairying than the lowlands. Beef herds are generally smaller and stocking rates of cattle therefore lower, which should favour Lapwings. But beef enterprises are much more often part of a mixed stock farm, with sheep, and the important point may be when cattle are turned out in spring. This may be up to a month later in upland areas than lowland and coincide more often with the trend to later nesting in upland grass (M. Shrubb, pers. obs.), resulting in desertion by birds which had established themselves in hitherto unstocked fields. It suggests that farms with cattle are less suitable for nesting lapwings irrespective of stocking rates due to the spreading of dung in spring.
Notes	Limitations identified by author	Fledging success has not been estimated as it is impossible to follow the fates of precocial chicks from nest,-record cards. Brood size has been calculated from the last observation of clutch size prior to hatching (very few counts of chicks in the nest or its immediate vicinity were

		recorded). This method may overestimate hatching success as eggs that fail to hatch or chicks that die during hatching may not always remain in the nest long enough for the observer to record on subsequent visits. Preponderance of nests identified in lowland grassland by CBC is attributed to differences in observer number and survey intensity.
	Limitations identified by review team	The review categorises grassland into rough grazing or improved. If the observer described the habitat as 'rough grass', 'open hill grazing', 'moorland' or 'marsh', or included in the habitat description such details as undrained boggy areas or infestations of rushes or thistles. Grass was classified as improved if it was described as 'ley', 'improved upland pasture', 'grass' or 'improved pasture/meadow', neither of these fit well with species rich unimproved meadows which are the focus of this review but the general principal of impacts of stocking rate trampling will apply similarly.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Use of BTO

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow priority habitat?
Study Citation	M. Shrubb (1990): Effects of agricultural change on nesting Lapwings Vanellus vanellus in England and Wales, Bird Study, 37:2, 115-127
Study Design Category	2
Assessed by & when	C.E. Pinches 23rd December 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Very well described.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes, employs the Common Birds Census monitoring programme and also factors in some additional comparative analysis which pre date this from 1940. It is suggested in the report that lowland grasslands may be over representing in the CBC
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Yes very well.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments: NA Correlative study

2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Yes comprehensive range of farm practices at critical nesting time assessed.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Comments: NA Correlative study.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: Fairly well as based on field observations.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Yes objective observations. .
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes employs CBC methodology.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Yes
3.4 Were outcomes relevant?	☐++	Comments: Yes

Where surrogate outcome measures were used, did they measure what they set out to measure?		
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: NA correlative study
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Yes records assessed over 40 year period.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NA	Comments: Not applicable
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: Yes in the sense that multiple factors effecting nesting success were observed.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Few details provided but it looks like simple statistics have been employed, i.e t test and are appropriate.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential	☐++	Comments: Yes.

confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Simpson, N.A. & Jefferson, R.G.
	Year	1996
	Aim of study	To conduct a comprehensive search of the agricultural and ecological literature for information relating to the use and impact of farm yard manure on the floristic composition of neutral grassland hay meadows, both unimproved and improved (MG3 -8) To establish current practice on hay meadows SSSIs where FYM is used via a questionnaire to English Nature's local teams (N.B this element of the report provided information for only 11/240 sites and can not therefore be seen to be representative.) To provide a brief summary report which will be used to guide best practice.
	Study design	3
	Quality score	++
	External validity	+
Population and setting	Source population	Species rich meadows of nature conservation interest NVC types MG3 - 8

	Eligible population	As above
	Inclusion and exclusion criteria	Assumptions that report makes are clearly set out, p.g 3 – 4.
	Setting	
Methods of allocation to intervention/control	Methods of allocation	NA
	Intervention description	NA
	Control/comparison description	NA
	Sample sizes	NA
	Baseline comparisons	NA
	Study sufficiently powered	NA
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	NA
	Secondary outcome measures	NA
	Follow-up periods	NA
	Methods of analysis	NA
Results		**Systematic review** Conclusions about the effects of FYM on the species composition of semi-natural meadows are largely subjective and not verified. However in general terms, as rates and frequency of application of FYM increase beyond a certain point (which varies according to background

		fertility) , there is a decrease in the richness and abundance of dicotyledonous herbs and an increase in competitive grasses resulting in an overall reduction is species richness and diversity. This is consistent with the effects of inorganic fertilisers on species rich grassland. Annual high rates (>30 t/ha of FYM cause scorching and bare patches to reseeded grassland . These rates applied annually or even less frequently would be very damaging, reducing species richness and diversity. FYM is a variable commodity with nutrient content dependent on many factors including handling, storage and weather conditions, Use of poorly rotted or inadequately composted manure on semi-natural meadows should be avoided, to avoid transfer of weed seeds, germination of which is much reduced by storage of 2 or more months. Where testing is practicable FYM should not be applied until the C:N ratio is less than 18:1. From available literature fresh cattle manure has a C:N ratio of 18 -26.4, whilst for composted manure reported values range from 11.7 – 15. In the absence of analysis storage for a minimum of 12 months is suggested. **Timing of FYM application** Verification is required of the amount of crop available N for different application timings for grassland ideally by soil type. Available evidence indicates that timing of dung application varied considerably from place to place from February - April to September to December. Both winter and spring applications of FYM allow opportunity for efficient utilisation subject to satisfactory soil conditions. Results of experimental studies Chambers 1994 looking at nitrate leaching losses on freely draining grassland soils, showed that manure type, application timing and over winter rainfall patterns all have a significant effect on leaching losses. **Application rates** Fertiliser experiments using FYM as at Park Grass and Palace Leas have applied relatively high rates **Park Grass experiment** 35 t ha annual application in Nov/Dec between 1856 and 1863Ceased due to smothering of

		herbage under dry climate. A more moderate 35t ha application 1 in 4 years has been applied in Plot 19 since 1905, **Palace Leas (Cockle Park)** Annual and alternate year dressings of 20 t per ha on annual and biannual cycles and 40 t ha per year. These represent extremely high additions and are not practically relevant for meadows of high nature conservation value. These levels of input may have been representative of certain fields on upland farms where the area available for mowing s limited to the more level and accessible fields and there is a need for large amounts of fodder to be conserved over long winters. **Periodicity of application of FYM** The authors conclude that periodicity (and rate) of FYM will influence yields and botanical composition of a meadow, if only on a cyclical basis. Further research is required. **Impact of FYM on floristic composition** FYM typically increases the amount and proportion of grasses in a sward at the expense of dicotyledous plants and lower plants (Park Grass;) Dodd et al. 1994 ascribed botanical data from the PGE plots to NVC communities and sub communities using match. For each treatment plot and time period the 3 NVC communities and sub communities which had the highest similarity coefficients were listed. The unlimed and limed PG receiving FYM were regularly matched MG3 ad MG5. They also regularly matched with the MG3a sub community (more species poor, grass dominated sub community and occasionally with MG6 and MG7 (although a lower coefficient values) suggest that FYM application rates at 35 t ha every 4th year may be sub optimal for the maintenance of species rich lowland meadows. Floristic change due to nutrient additions is thought to be caused by the following sequence – some species (usually grasses) generally grow faster and bigger than most other (mostly

		dicotyledonous) species when well fertilised. Competitors shade stress tolerator species, so the latter then grow less, reproduce less and there are less niches for these species, so resulting in change. The yield and composition of herbage and hay and the rate of change is dependent on considerable no of other factors
Notes	Limitations identified by author	Evidence gaps as set out below.
	Limitations identified by review team	Palace leas experiment may have been unjustifiably excluded, and may be relevant in the context of sustainable UHM management under higher rainfall more leaching. The view taken by the authors may be somewhat lowland centric.
	Evidence gaps and/pr recommendations for further research	Reassessment of nitrification rates and controlling factors in grassland. Effects of different periodicities (annual, triennial) application of FYM on different soils and under different rainfall regimes should be investigated. Better standardised recording and monitoring of FYM inputs, crop yields, management and botanical composition.
	Sources of funding	English Nature

Quality Assessment Checklist: Qualitative Study v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Simpson, N.A., Jefferson, R.G (1996). Use of farmyard manure on semi-natural (meadow) grassland., *English Nature Research Reports* (p. 97p.). [Peterborough].
Study Design Category	3
Assessed by & when	C.E. Pinches 24th November 2012

Section 1: Theoretical approach		
1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments: Yes, seeks to review available evidence on impacts of FYM on botanical composition of lowland meadows.
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments: Underpinning assumptions clearly set out,
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Defensible	Comments: Systematic literature review with clearly defined parameters search terms.

Section 2: Study Design		
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Defensible	Comments: Systematic literature review with clearly defined search parameters.

Section 3: Data Collection		
3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Appropriately ☐ Not Sure / inadequately reported	Comments: Not clear how references were searched for and whether this was systematic.

Section 4:Trustworthiness		
4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐ Not described	Comments: One of the authors, is an employee of the sponsoring agency, English Nature.

4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments:
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Reliable	Comments: Systematic review methods appear to be reliable. Note nothing could be drawn from the questionnaire aspect of this report as there were insufficient returns and data.

Section 5: Analyses		
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments: There is little data to analyses, instead the literature is reviewed and reported.
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:

5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	NA	Comments: NA literature review only
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Convincing	Comments: Findings convincing and reported unbiasedly.
5.5 Are the findings relevant to the aims of the study?	☐ Relevant	Comments:
5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	Comments: generally conclusions are reliable but in a couple of instances they are largely speculative and have been made in the absence of sufficient evidence.

Section 6: Ethics		
6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered? - Was the study approved by an ethics committee?	☐ Not Sure / not reported	Comments: NA

Section 7: Overall Assessment		
As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ ++	Comments: Well conducted.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Smith & Jones
	Year	1991
	Aim of study	To compare past and present practice in hay cutting times on meadows at five farms in the Yorkshire Dales and Cumbria, and assess current vegetation composition to determine if there is any association with sequence of cutting. To examine the phenology of common meadow species in one MG3 meadow and assess likely impact of changes in cutting date on these species.
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Upland hay meadows in the Yorkshire Dales and Cumbria
	Eligible population	Populations of upland hay meadows in Northern England

	Inclusion and exclusion criteria	Farms selected with help from the NFU who identified those farms for which detailed diaries had been kept.
	Setting	Historic start and finish dates for hay making were collected from 30 meadows across six farms in the Yorkshire Dales National Park and Ravenstonedale in Cumbria. Vegetation composition was subsequently sampled from these meadows. The phenological study took place at Bowberhead head meadow in Cumbria.
Methods of allocation to intervention/control	Methods of allocation	None required correlative study
	Intervention description	No intervention – correlative study investigating relationship between hay cutting start and finish dates and sward composition.
	Control/comparison description	No controls - correlative study
	Sample sizes	For the correlative study 30 meadows from which a total of 110 randomly located quadrats were recorded – the exact number of quadrats varied between meadows. Soil smaples down to a depth of 15 cm were also taken from each quadrat. For the phenological study – 5 randomly selected points were identified in the meadow and the developmental stage of the five closest individuals for each of 15 typical MG3 species was recorded together with the number of flowers present on each individual. Phenologies were constructed between 1st June and 21st August 1988. Data for Geranium pratense was collected from a nearby roadside verge. Records of flowering from grasses and composite species (Plantago lanceoloata and Sanguisorba officinalis) were based on visual estimates of the relative proportions pf the inflorescence in the 5 different distinct stages of flowering.
	Baseline comparisons	None recorded.
	Study sufficiently powered	Yes
Outcomes and methods of	Primary outcome measures	Cover/abundance of all vascular plant species present in the 1m2 quadrat using the DOMIN

analysis (inc effect size, CIs for each outcome and significance)		scale. Soil bulk density, pH, total carbon (LOI), available N and total N. Soil P, Ca. Mg, Na and K were also determined. Phenological study – for each hay meadow species the mean proportion of each flowering stage at each point in time was calculated over all 25 plants – providing an estimate of the species phenology.
	Secondary outcome measures	
	Follow-up periods	Hay cutting start and finish dates were assessed for meadows from 1947 to 1986.
	Methods of analysis	Canonical Ordination Analysis was used to determine the importance of the environmental variables of species composition, multicolinearity between the variables was reduced by removing some variables priori to each farm analysis. Significance was tested using the Monte Carlo test. The effects of other potentially confounding management variables, such as fertilizer use was controlled for by working out the degree of association between individual plant species phenologies and the sequence of cutting on each farm was assessed using the correlation coefficients plotted against the optimum date for ripe seed production as revealed by the Bowberhead phenological study.
Results		Between the years 1947 and 1986, hay cutting start dates showed little variation around the 1st July on the five farms studied. In contrast, hay cutting finish dates varied considerably with time, becoming far earlier in later decades as the time it takes to make hay significantly shortened, with the advent of mechanisation in the 1960s. Historic data indicate that pre-mechanisation the frequency of very late cutting was as regular as two in every five years on some farms. A significant relationship between sward composition and order of cutting was found on three of the six farms surveyed (*p* < *or* =0.03). However on the other three farms, where artificial fertiliser had been applied, this had the greatest effect on composition masking

		any effect of cutting order. The phenological study found that ripe seed are present at different times for different species. Red fescue *Festuca rubra,* cock's foot *Dactylis glomerata*, red clover *Trifolium pratense* and rough hawkbit *Leontodon hispidis*, produce seed from early August, whilst great burnet, *Sanguisorba officinalis*, knapweed *Centaurea nigra* and meadowsweet, *Filipendula ulmaria* have little ripe seed by 21st August. The authors suggest intermittent late cuts may be needed to enable adequate seed production and return for these species if early cuts are the norm.
Notes	Limitations identified by author	Any effects of the sequence of hay cutting on the vegetation are confounded by other management differences between the meadows as well as intrinsic site differences, for example outlying meadows on more infertile soils have a different range of species than those from more fertile meadows closer to the farm – often this reflects depth of the soil.
	Limitations identified by review team	Phenological aspect of this study is based on observations made on one site in one year so is likely to provide an imperfect representation of the phenology of other meadows, and in other years. The fact that Geranium sylvaticum population studied was from an adjacent road side verge may also reduce its representativeness as this population may not be adapted to as regular mid summer hay cutting.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Yorkshire Dales National Park Authority and University of Newcastle

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow priority habitat?
Study Citation	Smith, R.S., & Jones, L. (1991). The phenology of mesotrophic grassland in the Pennine Dales, Northern England – Historic hay cutting dates, vegetation variation and plant-species phenologies. *Journal of Applied Ecology, 28*, 42-59.
Study Design Category	2
Assessed by & when	CE Pinches 27th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Very well described.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes meadows represent species rich MG3 meadows. However the phenology of individuals of Geranium sylvaticum from a road side verge population may differ from that in an in meadow situation (being less adapted to mid –summer hay cutting) and therefore not be wholly representative of meadow populations of this species.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Yes for both the correlative and phenological aspects of the study.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison)		Comments: NA Correlative study

group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Yes, although this question relates better to the analysis in this instance.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Comments: NA Correlative study.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: The principle source of confounding factors relate to management variables other than cutting sequence (nutrient input) and also innate differences in soil depth across the meadows – fertility is well controlled for in the analysis.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Subjective (botanical % cover, DAFOR) and objective (soil nutrient status sampling) Significant efforts made to ensure QA etc.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Yes
3.4 Were outcomes relevant?		Comments: Yes

Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: NA
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Yes for the correlative study but the phenological study would have benefitted assessments over multiple years and/or multiple sites.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	Comments: Yes
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: Yes
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Yes
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of	☐+	Comments: Yes, the analysis of the impact of hay cutting dates on sward composition is well conducted taking into account the effects of differences in fertility.

bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design		The findings of the phenological study are interesting but are limited in the extent to which they can be translated across to other sites – as all measurements took place on one meadow in one year.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Smith & Rushton
	Year	1994
	Aim of study	To investigate vegetation changes arising when grazing by sheep and cattle is prevented for various periods of the year.
	Study design	1
	Quality score	++
	External validity	+
Population and setting	Source population	Upland hay meadows with MG3a (more species poor sub-community)
	Eligible population	North Pennine hay meadows
	Inclusion and exclusion criteria	
	Setting	Undergate meadow, Bowberhead, Ravenstonedale, Cumbria – Bowberhead and Piper Hole

		meadows SSSI
Methods of allocation to intervention/control	Methods of allocation	Random allocation of grazing treatments to randomly positioned plots within 3 Blocks
	Intervention description	1) No grazing at any time of year 2) Spring grazing only treatment (no grazing from the time of the haycut until Jan 1) 3) Autumn grazing only treatment (no grazing from January until the haycut) 4) Controls in which normal grazing regime was followed i.e 2 +3
	Control/comparison description	Yes, treatment 4 represents the normal regime.
	Sample sizes	Each treatment replicated three times Each year, 3 quadrats were randomly chosen from a central grid of 9 25cm x 25cm quadrats Samples were harvested 2-6 weeks in advance of the hay cut Mean dry weight of samples was calculated from 9 quadrats arising for each treatment
	Baseline comparisons	Vegetation sampled in first year 1987.
	Study sufficiently powered	Yes
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Mean dry weight of each species
	Secondary outcome measures	16 “species attribute” variables were constructed for each quadrat from data provided by Grime. For each quadrat for continuously varying species attribute variables such as nuclear DNA content were calculated using a weighted mean of biomass of all species. Categorical attributes were based on the weighted mean biomass of those species that were known to possess that particular attribute. Where species were intermediate in character for the attribute, their biomass was allocated to both categories.

	Follow-up periods	5 years
	Methods of analysis	Overall changes in species composition from 1987 to 1991 was assessed using DECORANA Detailed comparison of species of the species composition of the vegetation in the four grazing treatments was restricted to 1991 data. Species dry weights were log transformed. Species that occurred with an overall frequency of 20% were tested for their association with treatment by a one way ANOVA RDA was used to explore relationship between species and treatments, significance of relationship with treatment was tested using a Monte Carlo test. Associations between species attributes and the treatments was analysed by one factor ANOVA of the log transformed data. RDA was used to explore how these species attributes related to treatment. Changes in no and relative abundance of vascular plant species were assessed from a comparison of the species richness and diversity of the vegetation in each treatment in 1991 along with an assessment of change in vegetation community using MATCH.
Results		By 1991 all plots had diverged from the initial 1987 species composition, although the control and autumn grazed plots had remained very similar to each other throughout this period. Species composition in 1991 By 1991 the various grazing treatments had produced significant differences in the species composition if the vegetation. The most extreme response was elicited by the complete cessation of grazing, which was the only treatment for ewhich there was a significant change in the number of species (25% decrease). A distinct group of species, particularly grasses Bromus hordaceus, festuca rubra, Alopecurus pratensis , DActylis glomerata and Holcus lanaatus became dominant under this

		treatment. Varying the time of the grazing between autumn and spring favoured different groups of species. Autumn grazing alone favoured the grasses Anthoxanthum odoratum, Lolium perenne, Poa trivialis and Cynosurus cristatus. Spring grazing alone favoured the herbs geranium sylvaticum, Cirsium heterophyllum and Sanguisorba officinalis – changes in composition reflect changes in the relative abundance of these species. Grazing in the spring and autumn was essential for the maintenance of Trifolium repens. From 1991 data species richness decreased as the sum of the standing crop of all species increased, this relationship did not match Grime's hump backed model. When comparison of species richness and harvested biomass was analysed over the enture time series of the experiment and restricted to data from the control only – the data did fit the hump backed model indicating that species richness in the meadow was in a dynamic equilibrium with harvested biomass on an annual basis, potentially in response to climate. Four species attributes were significantly associated with the treatments, relating to reg High nuclear DNA has been linked with a species ability to grow early in the year, when temperatures atre suboptimal for growth. Indeed as expected species with a high nuclear DNA content were associated with the un-grazed treatment. Findings emphasise the importance of regeneration niches in the maintenance of high species richness. Bare soil, created by cows hooves when meadows are grazed in autumn creates gaps large enough for seedling establishment, in the study competitive ruderals with autumn germinating seeds were favoured in the autumn grazed and control treatment
Notes	Limitations identified by author	A subsidiary trend of change through time irrespective of treatment was identified – for which several reasons are suggested but cause is unknown. Between 1988 – 1990 there were a series of very dry summers which may contribute to the change. The destructive sampling strategy may have had an impact, as potentially indicated by the disappearance of Rhinanthus minor from stands.

	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Small ecological grant from the British Ecological Society and the Yorkshire Dales National Park Authority

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Smith, R.S., & Rushton, S.P. (1994). The effects of grazing management on the vegetation of mesotrophic (meadow) grassland in Northern England. *Journal of Applied Ecology, 31*, 13-24.
Study Design Category	1
Assessed by & when	CE Pinches, 13th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Yes, though details of exact species composition not provided.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes, although representative of the less diverse sub community of MG3.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	Subjective selection of field site, no reasons provide.

Section 2: method of allocation to intervention(or comparison)

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: The experiment employs a fully randomised block design, with three replicate blocks at each site.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: All treatments including control are well described and would enable replication. Comparisons are appropriate.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Yes, no problems
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes subject to hay cut across entire emperimental site
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes representative of upland hay meadows
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Yes differing grazing regime reflect s different practices

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable?	☐++	Comments: Objective - Mean dry weight of each species

Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?		16 “species attribute” variables were constructed for each quadrat from data provided by Grime. For each quadrat for continuously varying species attribute variables such as nuclear DNA content were calculated using a weighted mean of biomass of all species.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Yes, derived attribute variables appropriate
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Probably.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: Yes
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted	☐++	Comments: Yes

standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Species that occurred with an overall frequency of 20% were tested for their association with treatment by a one way ANOVA RDA was used to explore relationship between species and treatments, significance of relationship with treatment was tested using a Monte Carlo test. Associations between species attributes and the treatments was analysed by one factor ANOVA of the log transformed data. RDA was used to explore how these species attributes related to treatment. Changes in no and relative abundance of vascular plant species were assessed from a comparison of the species richness and diversity of the vegetation in each treatment in 1991 along with an assessment of change in vegetation community using MATCH.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Yes.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: Yes A subsidiary trend of change through time irrespective of treatment was identified – for which several reasons are suggested but cause is unknown. Between 1988 – 1990 there were a series of very dry summers which may contribute to the change. The destructive sampling strategy may have had an impact, as potentially indicated by the disappearance of Rhinanthus minor from stands.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and		Comments: Yes, the findings can be extrapolated to other MG3 meadows

nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What management regime maintains the diversity of flora and fauna of the upland hay meadow priority habitat?

Study details	Authors	Smith, R.S., Buckingham, H., Bullard, M.J., Shiel, R.S., & Younger, A. (1996). The conservation management of mesotrophic (meadow) grassland in northern England. 1. Effects of grazing, cutting date and fertilizer on the vegetation of a traditionally managed sward. *Grass and Forage Science, 51*, 278-291.
	Year	1996
	Aim of study	To determine the interacting effects of different grazing, fertilizer and cutting date treatments on the vegetation of an upland hay meadow
	Study design	1
	Quality score	++ (for fertilizer and hay cutting aspects) + (for grazing)
	External validity	+
Population and setting	Source population	North Pennines Upland Hay Meadows
	Eligible population	MG3b species rich Upland Hay Meadow

	Inclusion and exclusion criteria	-
	Setting	Gillet Farm, Upper Teesdale, Co Durham
Methods of allocation to intervention/control	Methods of allocation	Partially randomised due to need to control grazing livestock – Grazing randomly allocated to 3 blocks, of 3 plots each, hay cutting treatment applied to 3 sub plots within plots and fertiliser treatment randomly allocated to two sub-sub plots within each grazing treatment.
	Intervention description	At plot scale - Grazing – a) no grazing b)autumn cattle grazing September to October stocking rate 0.54 ha-1 c)autumn grazing plus spring grazing for 1 week in early/mid May with sheep. Plots divided into 3 sub plots - Hay cutting dates – a) 14th June b)21st July c) 1st September (Cut herbage turned once and dried on the sub plot prior to its removal). Sub plots divided into two sub-sub plots and two fertilizer treatments applied. 1) No fertilizer or 2) 80kg ha N plus 40kg ha P and K, applied in mid April in each year.
	Control/comparison description	Control represents the traditional management regime, ie. No mineral fert, 21st July cutting date for hay, autumn grazing with cattle and spring grazing with sheep. NB Previous management regime on the experimental site did not include spring grazing and did include annual application of light dressing of FYM.
	Sample sizes	3 x replication.
	Baseline comparisons	Between treatment comparison only no baseline data pre experiment collected.
	Study sufficiently powered	Ok for fertiliser and cutting date but grazing treatment may have insufficient power to detect sig effects (few d of f).
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	Species presence and % cover of each species recorded in 5 randomly placed 0.0625 m2 quadrats in each sub-sub plot in June and July 1990, 1991 and 1992. In June and July 1993 data as above collected but within 2 randomly placed quadrats per sub

significance)		plot each within nested set of 0.625 m2, 1 m2 and 4m2.
	Secondary outcome measures	16 “species attribute” variables were constructed for each quadrat from data provided by Grime. For each quadrat for continuously varying species attribute variables such as nuclear DNA content were calculated using a weighted mean of % cover of all species. Categorical attributes were based on the weighted mean of the sum of the % cover of all species that were known to possess that particular attribute. Where species were intermediate in character for the attribute, their cover was weighted by 0.5.
	Follow-up periods	1989 - 1993
	Methods of analysis	Vegetation change over 4 years Overall species change from 1990 to 1994 was assessed by Detrended correspondence analysis. Treatment effects Redundancy Analyses were used to 1) relate species to each other and to treatment 2) relate “species attributes” to treatments and to species. ANOVAs on species attributes and on 11 species that occurred with an overall frequency of greater than 80%. Change in wider vegetation context was made by categorised full nested quadrats data from 1993 by Tablefit and comparing treatments. Comparison of associatiob of NVC type with treatment was tested for by Chi square test.
Results		**Individual treatment effects at experimental end point 1993**

		Addition of fertilizer significantly reduced the number of species by between 12 and 21% (P<0.001). Cutting date and grazing had no such significant effects, with the exception that number of species at 0.0625m2 was significantly higher (p<0.05) under autumn and spring grazing. Fertiliser use lead to a significant increase in the abundance of species that were competitors, and/or that are able to rapidly capitalise on increased nutrient availability, namely those capable of vegetative reproduction (e.g grasses) with seed 1) capable of immediate germination 2) mainly germinating in autumn but maintaining a small persistent seed bank (Type III). In contrast, when fertilizer was not applied, there was a significant increase in the abundance of species that were stress tolerators, ruderals and stress tolerating ruderals and/or that had large persistent seed banks (Type IV), had seed that took longer to germinated or required chilling to promote germination. Ruderal species were significantly (p<0.001) more abundant with successively earlier hay cuts. Whilst stress tolerating ruderals were favoured by the 21st July hay cut (p<0.05). The latest hay cut (1st September) was significantly associated with increasing abundance of competitor species (p<0.001) with, light or large and persistent seeds (Type IV) p <0.01 and vegetative spread p<0.05). **Interactions between treatments** The most important interactions between treatments were with the combined effects of late cutting date and use of fertilizer. No of species was most reduced when both fertilizer and 1st September cutting date treatments were applied and these treatments was associated with the NVC types U4b and MG7c, most divergent from MG3b. This combination of treatments encouraged competitive species principally grasses, which can spread vegetatively, germinate rapidly and have persistent seed banks. Under the traditional management regime (i.e. no fertilizer, 21st July cutting date, grazing

		autumn and spring) species composition remained relatively static. Complete cessation of grazing caused considerable changes. Diversity within traditional management regimes seems to be controlled through the provision of niches for slower growing species and those that require germination niches for their larger seeds that take longer to germinate or need special treatment such as chilling to break dormancy. MG8 was associated with some of the early and late cutting date treatments.
Notes	Limitations identified by author	Potentially insufficient power for grazing component of experiment Interactions between treatments may become significant over a longer time period. Late cutting, no grazing treatment would not normally be applied consistently over 4 years in a farming system
	Limitations identified by review team	Study doesn't account for impact of normal light dressing of FYM on sward and it's interaction with different grazing cutting regime.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Agricultural and Food Research Council, EN and MAFF

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review___________________________

Name of Review Sub-topic (if any): __Hay Meadows_____________________________

Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?
Study Citation	Smith, R.S., Buckingham, H., Bullard, M.J., Shiel, R.S., & Younger, A. (1996). The conservation management of mesotrophic (meadow) grassland in northern England. 1. Effects of grazing, cutting date and fertilizer on the vegetation of a traditionally managed sward. *Grass and Forage Science, 51*, 278-291.
Study Design Category	1
Assessed by & when	CE Pinches, 19th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐ NR	Comments: Not reported in any detail/.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Manner by which experimental site was selected is not described but is not random. Site is representative of MG3b species rich Upland Hay Meadow sub community.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+ or NR	Comments: No information is provided on how plot location was determined within the meadow.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: The experiment employs a partially randomised designed with three blocks of 3 plots. Three grazing treatments were randomly allocated to the 3 blocks. Each plot was sub divided 3 sub plots and 3 hay cutting date treatments randomly allocated between them.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Yes well described - experimental treatments were designed to compare traditional management with quite extreme variants, namely modern variants of this and exceptional historic variants. For e.g whilst late September cutting may occur perhaps 1 or 2 in years in 5 before mechanisation they would not have been implemented consistently over 4 years.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes Fertiliser rates were considerably higher in the plus fertilizer treatment than rates allowed in the guidelines. Plus continued implementation of late hay cut and no grazing not very representative.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: No contamination reported. Whole meadow had previously received same management which included annual light dressing of FYM and no spring grazing.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Not reported so assumed not.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Yes
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Yes reflects traditional management of MG3

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Both - Subjective botanical assessments - % cover of each species present, presence from randomly positioned quadrats.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Yes, derived species attribute values also relevant.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Probably, though authors note that interactions between treatments have taken longer to attain significance in similar studies on lowland grasslands. .

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Differences in initial composition of blocks apparently not tested or controlled for.

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	Comments: No power analysis conducted but there is suitable replication of treatments for fertilizer and cutting date treatments, maybe insufficient for grazing.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	Comments: Yes, but may have been useful to control for any starting differences in botanical composition especially given low no of d of f for grazing treatment which renders blocking useless for this treatments.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: Yes, with possible caveat that grazing element was insufficiently powered. Differences in initial botanical composition between blocks not accounted for Grazing element should be scored 1+ whilst all other aspects fertilizer and cutting can be scored 1++
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Yes, the findings can be extrapolated to MG3 a similar management history.

Evidence Table

Name of Evidence Review:	Upland Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	R.S. Smith1, R.S. Shiel1, D.Millward2, J. Simkin3 and S. Pratt4
	Year	2012
	Aim of study	To determine how the duration and intensity of spring grazing affect hay yield and quality, plant species diversity and composition, and the performance of key community character species. To consider the ecological mechanisms underlying plant growth and development in the experiment, particularly the link with spring temperature.
	Study design	1
	Quality score	1++, EV + or ++
	External validity	
Population and setting	Source population	MG3b meadow, Wensleydale
	Eligible population	Species rich MG3 meadows
	Inclusion and exclusion criteria	

	Setting	Unimproved MG3b Meadows (Burrey and 3 acre) Thornton Rust, Wensleydale
Methods of allocation to intervention/control	Methods of allocation	Fully factorial randomised experimental design
	Intervention description	All combinations of four shut up dates (1st February, 1st May, 15th May and 27th May) and two grazing intensities (sward heights of (high intensity)3-4cm and (low intensity) 5-6cm)
	Control/comparison description	The 15th May shut up date represents the traditional meadow shut up date for the area for comparison with the other dates.
	Sample sizes	3 blocks, within which 8 plots – 3 x replication of each treatment combination
	Baseline comparisons	Effect of baseline composition controlled for by using data 2008 in ANOVA with years 2008 – 2011 as repeated measures and by blocking.
	Study sufficiently powered	Yes.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Plant species presence within 1 m2 quadrats and % cover. (data amalgamated to provide a species list for each plot (species 3m2) with mean % cover abundance values) Sward height and height of key community character species from February to June (mean of closest individual plants from 5 random located positions in the plots) On 3 occasions from mid June to mid-July 2008 -11, a phenological survey was undertaken using random sampling positions to determine a record of number of flower buds, open flowers, unripe and ripe seed capsules and dehisced seed capsules Standing crop and hay quality assessed from 5 x 0.0625m2 random quadrats per plot prior to the main hay cut.

		Soil samples taken in 2008.
	Secondary outcome measures	Species diversity (using Shannon and Simpson indices) and also Ellenberg fertility.
	Follow-up periods	Experiment is in year 4 of 5, 2008 – 2012, final report will be provided in 2013.
	Methods of analysis	**Impact of temperature on growth** Association between spring temperatures (accumulated temperatures from 1st Jan) and plant growth (mean sward height and height of individual species) was investigated as one of the underlying ecological mechanisms affecting vegetation response to the applied treatments. **Treatment effects** Multifactorial ANOVA used to assess effects of block, grazing intensity and shut up date and interactions on a range of vegetation and crop characteristics, for 2011 data alone and for 2008 – 2011. In addition a repeated measures ANOVA tested differences between year. Ordination (RDA) was used to provide a review of the main trends in the experiment in the first 4 years, using blocks as co-variables. Tablefit used to provide a similarity coefficient between the species composition of each plot and standard floristic table for the community. Assessment of mean characteristics by treatment and block were also assessed within national context of similar grasslands.
Results		**Shut up date** Over the 4 years of the experiment , the 1st Feb shut up date produced grass swards with a significantly greater similarity to MG3b vegetation ($p=0.003$) than swards from the later shut dates. The latest shut up date significantly reduced Simpson

		(p=0.001)and Shannon diversity (p<0.001) and Shannon evenness (p=0.03) The later the shut up date the more delayed the initiation of flowering in key community character species (hay rattle, red clover, pignut) and therefore delayed seeding. **Grazing intensity** Over 4 years more intense grazing significantly reduced MG3b similarity (p=0.011), Simpson (p= 0.028) and Shannon diversity (p = 0.01) and increased Ellenberg fertility (=0.039). **Interaction effects** Species richness progressively decreased with later shut up date at the high grazing intensity, so the earlier shut up date had significantly more species than the last (p=0.017) Effect on yield Later shut dates and higher grazing intensity significantly decreased the yield of hay (p<0.002) but significantly increased its quality (p<0.001), when the latter is defined by N content and digestibility **Ordination** Earlier shut dates and lower grazing intensity were correlated with increasing standing crop, plant diversity and similarity to MG3b grassland and with decreasing Ellenberg fertility. Changes in shut up date and grazing intensity altered the balance between species rather than altering species richness. **Interaction of management with climate** Under normal farming practices, i.e. shut up at 15th May and with increased growth due to warmer springs in the period since 1990, sheep will have eaten much of the vegetation growth and in doing so may adversely affect plant species that would still have been dormant under colder springs.

		This study provides an indication of this, in finding that by late May the key community character species, in particular hay rattle but also pignut, wood cranesbill, red clover and greater burnet were all taller under the earlier shut up date and low intensity grazing treatment. Late removal of sheep gives less time for the development of ripe seed and dehisced capsules for key species by mid July, i.e. by the hay cut. In summary, delaying the date at which sheep are removed for the growth of the hay crop until late May delays maturation of the sward (by constantly promoting new leaf growth and not encouraging development of flower stems, flowers and seeds) and appears to drive a reduction in hay rattle populations. Early shut up dates gives species that flower late in spring a head start and allow them to maximise seed production. **Interim conclusion** Maintenance of MG3 grassland would be facilitated by low intensity spring grazing but particularly the earlier removal of sheep for the hay crop. This may require a fundamental shift in starting date of the lambing season with it being brought forward to enable earlier shut dates and growth of the hay crop. This in turn would probably allow for an earlier hay cut than the existing management schemes permit.
Notes	Limitations identified by author	Species composition of swards were surveyed in systematic sequence following shut up date (across a 3 – 4 period) justified by the need for accurate species identification. For this reason the effect of temperature on species composition within each year were impossible to recognise due to the confounding effects of recording sequence. This was controlled for in the ordinations by using sampling date as a co-variable. Although results suggest that shut up date rather than sampling sequence is the main factor this will be tested in 2012 by gathering real data.
	Limitations identified by review team	Year by year differences were not consistent, for example the effect of different shut dates on vegetation character disappeared completely in 2010. The experiment would benefit from an extension to ensure overall trends persist.
	Evidence gaps and/pr	

	recommendations for further research	
	Sources of funding	DEFRA

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ____Upland Evidence Review

Name of Review Sub-topic (if any): ______Upland Hay Meadows________________________

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows
Study Citation	Smith et al (2012) SPRING GRAZING IN NORTHERN HAY MEADOWS: INFLUENCE OF THE TIMING AND INTENSITY OF SHEEP GRAZING ON THE FLORISTIC DIVERSITY AND RESTORATIVE POTENTIAL. The Northern Hay Meadow Project BD1467 2011-12 Annual report to DEFRA 27 February 2012
Study Design Category	1
Assessed by & when	CE Pinches, 17th October 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Yes, MG3b meadows at Thornton Rust well described though detail of botanical composition, soil characteristics is not provided .
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: Yes. When the study site was examined in the context of MG3 swards studies in a wider Defra study it was shown to be a typical example of a traditionally managed diverse sward.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described?	☐++	Comments: Variation in vegetation at experimental site controlled for by blocking (the 3 blocks did differ in their species richness). Three 1m2 quadrats randomly positioned within each plot, excluding a 1m wide boundary.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Fully randomised allocation
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Yes, very well described.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: There were a few slight deviations from the experimental design, namely: In 2008 the February 1 shut up date was applied on 1 April when the project started. Sward height for high intensity was raised from 2-3cm to 3-4cm after 2009. Due to the growing season not starting until early March it was impossible to create different sward heights prior to the 1st Feb shut up date. As a consequence “low intensity grazing” for this shut date was defined as “no spring grazing” whilst the “high intensity grazing” comprised of one off intensive grazing to reduce the sward to 3-4cm in mid April.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐+	Comments: See above deviations and justifications for amendments to 1st Feb shut up treatments well described.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the	☐++	Comments: Yes

England/UK Resource.		
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Yes shut up dates were chosen to represent 1) traditional meadow shut up date (15th May), 2) a later date to extend the grazing period and 3) an earlier date (1st May) thought to represent the start of the growing season in the upper reaches of the dales 4) an earlier date (1st Feb) to give complete freedom from spring grazing.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Subjective assessment of vegetation composition within three 1m2 quadrats together with objective assessment of a number of other variables, sward height, phenological variables, standing crop and hay quality.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes, assessments made of both ecological and agronomic outcomes.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++ ☐NA	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes, although the significance of treatment effects across years was inconsistent. Ideally these effects would be tested over a longer time period.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Yes, the final years data from the 2012 field season have yet to be included.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Significant effects of blocks were picked up in the analyses, but since allocation of the treatments was fully randomised this variation was correctly attributed.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	Comments: Sample size and degree of replication satisfactory
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes.
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Yes, comprehensive and well considered analysis
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: Yes
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?		Comments: Yes, though note unusual weather 2011 was particularly warm and dry in April whilst 2008 2010 were very cold in March and April.

Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat? Sub questions a and c

Study details	Authors	Smith R.S. , Pullan, S. & Shiel, R.S.
	Year	1996
	Aim of study	To quantify the amount of seed shed through hay making under different cutting dates.
	Study design	1
	Quality score	1 +
	External validity	EV +
Population and setting	Source population	MG3 grassland
	Eligible population	MG3 meadows with Pennine Dales ESA
	Inclusion and exclusion criteria	
	Setting	Gillet Farm, Upper Teesdale, Northumberland
Methods of allocation	Methods of allocation	Allocation of 3 grazing treatments was randomly within the 3 paddocks in each blocks.

to intervention/control		Each paddock was then divided into 3 sub plots and 3 hay dates were randomly allocated. Each sub lot was then divided into two sub, sub plots with 2 fertiliser treatments randomly allocated
	Intervention description	Grazing: 1) no grazing 2) autumn grazing with beef cattle and calves 3) autumn grazing as in 2 plus spring grazing for a week in May with ewes. Hay cutting dates 1)14th June 2) 21st July 3) 1st September Fertiliser 1) no fertiliser 2) 80kg ha-1 N plus 40kg ha-1 phosphorus and potassium
	Control/comparison description	Combinations of the above interventions treatments were chosen to mimic the traditional management regime (no use of mineral fertiliser, 21st July hay cut, autumn grazing with cattle, spring grazing with sheep) a modern variant of this (use of mineral fertiliser, 14th June hay cut, autumn grazing with cattle, spring grazing with sheep) and exceptional historic variations (1st September hay cut, and no spring grazing). No specific control, i.e. continuation of past management was provided but the experimental design allows each factor in turn to be examined
	Sample sizes	Vegetation and seed sampled from 162 quadrats across the treatments, n values for each treatment not provided. Hay cutting simulated through destructive harvesting of vegetation within in plot quadrats, vegetation then dried and shaken repeatedly to mimic drying and tedding of hay in the field.
	Baseline comparisons	Seed fall from traditionally managed hay making assessed outside the experimental plots via seed collection from 20 randomly positioned circular pitfall traps . However as hay was cut on 7th July , there is a 14 day gap to the hay cut date of the 21st July simulated hay cutting treatment.
	Study sufficiently powered	Sampling and degree of replication of treatments is sufficient.

Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation was destructively sampled. % Species composition of sward from biomass of each plant species Seeds shed identified to species level and numbers counted.
	Secondary outcome measures	
	Follow-up periods	Experimental management regime set up in 1989, with sampling of seed shed taking place only in 1990.
	Methods of analysis	ANOVA to determine treatment effects on mean number of seeds by species and overall. Student's t test used to compare the mean seed data by species from pit fall traps seed (traditionally managed hay crop cut on 7th July) with total number of seeds from the simulated 21st July cut.
Results		Cutting date: A significant effect of cutting date was found for 17 of the 23 species analyses and for all the all grass seed and all forb seed categories. Overall the 14th June hay cut shed relatively small amounts of see (15% of the quantity shed on 1st September), with slightly more seed contributed by forbs. The traditional hay time (21st July) shed more seed (34% of the quantity on 1st September) with nearly equal amounts of forb and grass seed. **Cutting on or around this date therefore favours the return of seeds for many , though not all forb species.** The delayed hay cut (1st September) shed the greatest quantity of seed but this was heavily biased towards grasses. • There was a highly significant increase in grass seeds significant decrease in forb

		seeds with later cut date ($p<0.001$). • Individual species behaved differently but the main trend in seed production was for forbs to shed seed generally during June and July and grasses during August and September. Grazing: Only 4 species showed significant responses and there was significant effect on grass or forb seed overall, sig diff were: • significantly more seed ($p<0.01$) of Holcus lanatus under the no grazing treatment ; significantly more seed ($p<0.001$) of Trisetum flavescens under autumn grazing; • significantly more seed of Ranunculus repens ($p<0.05$) and Helictotrichon pubescens with autumn and spring grazing. Fertilizer: • There was significantly ($p<0.001$) more grass seed and significantly less ($p<0.001$) forb seed when fertiliser was applied. There was: • Significantly less seed of Plantago lanceolata, Rhinanthis minor and Bellis perennis when fertiliser was applied and significantly more seed of Helictotrichon pubescens. Comparison with in field hay making: • With the exception of Bellis perennis, more seed was obtained from the pitfall traps than by the simulated hay making. Conclusions The restricted working (turning) of the grass crop associated with silage making may reduce the quantity of seed that is returned to the soil, leading to particular decreases in the populations of those species which rely on regeneration from seed, particularly

		those that produce small amounts or are rare in the sward. In addition, earlier cutting dates will reduce the amount of seed shed.
Notes	Limitations identified by author	Since sampling only took place in 1990, results are representative of conditions in this year. No assessment of seed viability was made. There was likely to have been an edge effect in the small quadrats sampled with seeds from the surrounding vegetation dispersing into the sample.
	Limitations identified by review team	14 day discrepancy in cut dates from infield hay cutting and closest simulate treatment date which made the comparison between the two less valid. No multifactorial results presented which would have enable relative statistical importance of treatments to be assessed
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Agricultural and Food Research Council and British Ecological Society.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ___Uplands Evidence Review____________________________

Name of Review Sub-topic (if any): Upland Hay Meadows_______________________________

Review Question	What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?
Study Citation	SMITH, R. S., PULLAN, S. & SHIEL, R. S. 1996. Seed shed in the making of hay from mesotrophic grassland in a field in Northern England: effects of hay cut date, grazing and fertilizer in a split-split-plot experiment. *Journal of Applied Ecology,* 33, 833-841.
Study Design Category	1
Assessed by & when	CE Pinches, 17th October 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	Comments: Yes, vegetation community, soil and climatic properties well described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Yes, although no details provided on basis for experimental site selection. Presumed to be on basis that site is representative of wider MG3 grasslands present across Northern Pennines. The historic management of the experimental site differed only from traditional practice in its lack of spring grazing by sheep.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: Yes, experiment employs a fully randomised block design. Blocks were aligned across the south facing slope so controlling for potential variation attributable to this feature.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Yes fully randomised.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Yes very well described , repeatable and appropriate. Comparison of specific management of study site was via controlled sampling of seed shed during actual hay cutting.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Yes, adequate. Only potential source of bias is degree and intensity of grazing, since livestock had access to wider paddock. However, efforts were made to ensure exposure was consistent by ensuring sward was grazed to uniform height of approximately 3 cm.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: Sub, sub plots for fertilizer treatment measure 10 by 10 m, no mention is given to discard areas so even with hand application of fertiliser there may have been some transfer of nutrients to no fertilizer treatments, via runoff for example.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR	Comments: None reported.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Botanical composition and past management are typical for MG3.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Treatments were chosen to mimic the traditional management regime (no use of mineral fertiliser, 14th June hay cut, autumn grazing with cattle, spring grazing with sheep), a modern variant of this (use of mineral fertiliser, 21st July hay cut, autumn grazing with cattle, spring grazing with sheep) and

		exceptional historic variations (1st September hay cut, and no spring grazing).

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Objective, destructive harvest of grass crop with species composition, species biomass and seed composition and number assessed.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐-	Comments: No, outcomes were measured only once, a year after the experimental treatments were applied and are insufficient to pick up long term effects in particular of grazing regime and fertiliser.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups	☐++	Comments: No, all ok.

in important confounders at baseline?		
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	Comments: Sample size adequate.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes, effect size presented with respect to actual % species composition/cover and % frequency at which seed was found. No standard errors were provide on actual no of seeds as analyses were conducted on transformed data.
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	Comments: Yes methods appropriate, data transformed where necessary. No multifactorial results presented which would have enable relative statistical importance of treatments to be assessed.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Yes, P values provided.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐+	Comments: Good experimental design
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Yes but note that results are unique to growing conditions in 1989/1990 care needs to be taken in ensuring climatic conditions then are representative of climate over longer timescale.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Smolders, A. J. P, Lucassen, E. C. H. E. T., van der Aalst, M., Lamers, L. P. M., Roelofs, J. G. M. **Year:** 2008	**Source population:** The relevant part of the manuscript was focused on ex-situ experiments using ex-arable soil **Setting:** Ex-arable soil in the Netherlands	**Methods of allocation:** ex-situ studies **Intervention description:** Liming of 0 (control), 10 (expt 1) or 5 (expt 2), and 20 g per kg soil	**Primary outcome measures:** 1st experiment available P. 2nd experiment growth of Juncus **Secondary outcome measures:** N/A	1st experiment shows that available phosphorus is reduced after applications of lime. The 2nd experiment backs this up, but does not show a corresponding decrease in Juncus growth. Since there are only two replicates per treatment, though, this result is not reliable	**Limitations identified by author:** None **Limitations identified by review team:** Underpowered

Aim of study: To find the effect of liming on Juncus growth **Study design:** Part observational study, part non-randomised control trial **Quality Score:** - **External validity:** - **Overall score: -**		**Control / comparison description:** No lime added **Sample sizes:** **Baseline comparisons:** N/A **Study sufficiently powered:** In 1st experiment 8 replicates per treatment which seem sufficient. In 2nd experiment 2 replicates per treatment, certainly insufficient	**Follow-up periods:** 3 months, which is meaningful, but a longer period would be useful **Methods of analysis:** No statistical analysis - only means, with confidence intervals for 1st experiment. Not for 2nd, because only 2 replicates per treatment		**Evidence gaps and recommendations for further research:** **Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	b - methods of controlling rushes
Study Citation	Smolders, A. J. P, Lucassen, C. H. E. T., van der Aalst, M., Lamers, L. P. M., Roelofs, J. G. M. (2008). Decreasing the abundance of Juncus effusus on former agricultural lands with
Study Design Category	Part observational study, part non-randomised control trial
Assessed by & when	Kate Fagan, 9th November 2012

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	The experiments utilised soil collected from ex-arable areas. The P study was based on ex-arable and reference sites - very little detail given about the reference sites

1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐-	No details given to demonstrate whether they are representative or not
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	As above - no details

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	NR	No information about this, but much less important than in some experiments since all ex-situ
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Yes
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate?	☐+	Comments:

Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)		
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Unlikely to have been any problems
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Unlikely
2.6 Were the	☐++	Ex-arable

wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.		
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	NR	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments:

3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:

3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Yes, it was meaningful, but a longer timescale - taking in reproductive success - would have been useful

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Soil was sampled from one site and pooled before dividing and adding different proportions of lime for the different groups
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard.	☐-	Almost no statistical analysis is mentioned, but sample sizes seem sufficiently large for experiment one. For experiment two there were only two repeats of each treatment, certainly insufficient

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		
4.3 Were the estimates of effect size given or calculable?	☐-	It doesn't appear that statistical analysis was employed

4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐-	Although statistical analysis was used when investigating P levels and Juncus cover in field soil, it doesn't appear that any statistical analysis was carried out for the two experiments. Confidence intervals were demonstrated in one figure for the 1st experiment (effect of liming on available P) but no CEs were given for the second experiment (investigating Juncus growth on limed/non-limed field soils)
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐+	It would have been more useful if the pH values had been given before and after - without those, the experiments are less meaningful

5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Despite the problems with analysis, there don't appear to be significant problems with the experimental design, and it is clear that Juncus growth is not diminished by lime addition
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	This experiment doesn't relate to the uplands, but in this question the results should still be relevant

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Year: **Aim of study:**	**Source population:** **Setting:**	**Methods of allocation:** **Intervention description:**	**Primary outcome measures:** **Secondary outcome measures:**		**Limitations identified by author:** Limitations identified by review team: Failure to control for effect of baseline vegetation composition of plots in detailed comparison of species composition and species attributes between treatments in 1991. Baseline vegetation shoudl have been treated as a covariate.

Study design: **Quality Score:** **External validity:** **Overall score:**		**Control / comparison description:** **Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Follow-up periods:** **Methods of analysis:**		**Evidence gaps and recommendations for further research:** **Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias?	☐++	Comments:

Were the inclusion / exclusion criteria explicit and appropriate?		

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments:
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments:

2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments:
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Comments:

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments:
3.2 Were all outcome measurements complete?	☐++	Comments:

Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant?		Comments:

If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments:

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	**Comments**:
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	**Comments**:

4.3 Were the estimates of effect size given or calculable?	☐+	**Comments**:
4.4 Were the analytical methods appropriate?	☐+	**Comments**:

Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?		
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐+	**Comments**:
[redacted]		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐+	Comments:
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?	☐+	Comments:

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?		

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Gruebler, M. U.; Schuler, H.; Horch, P; Spaar, R.	**Source population:**	**Methods of allocation:**	**Primary outcome measures:**		**Limitations identified by author:**
Year:	**Eligible population inclusion & exclusion criteria:**	**Intervention description:**	**Secondary outcome measures:**		**Limitations identified by review team:**

Aim of study:					**Evidence gaps and recommendations for further research:**
Study design:	**Setting:**	**Control / comparison description:**	**Follow-up periods:**		

Quality Score: **External validity:**		**Sample sizes:** **Baseline comparisons:** **Study sufficiently powered:**	**Methods of analysis:**		**Sources of funding:**

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	
Study Citation	
Study Design Category	2
Assessed by & when	

1.1 **Are the source population(s) or area(s) well described?**		Comments:

e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐-	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NR	
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	
2.4 How well were likely confounding factors identified and controlled?		

Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	
2.5 Is the setting applicable to the UK?	☐+	

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐-	Comments:
3.2 Were all outcome measurements complete?		

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐-	
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	

3.6 Was the post-treatment time interval meaningful? Was the follow up long enough to assess long-term effects?	☐+	Comments:

4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	

4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	

5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	

Study Details	**Population and setting**	**Methods of allocation to intervention / control**	**Outcomes and methods of analysis** (inc effect size, CIs for each outcome and significance	**Results**	**Notes**
Authors: Year: Aim of study: Study design: Quality Score: External validity: Overall score:	Source population: Setting:	Methods of allocation: Intervention description: Control / comparison description: Baseline comparisons: Study sufficiently powered:	Primary outcome measures: Secondary outcome measures: Follow-up periods: Methods of analysis:		Limitations identified by author: Limitations identified by review team: Evidence gaps and recommendations for further research: Sources of funding:

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any):

Review Question	
Study Citation	
Study Design Category	
Assessed by & when	

1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments:
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions?	☐ Clear	Comments:

-is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?		
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	Comments:
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question?		Comments:

-Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure	

3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Not Sure / inadequately reported	Comments:

4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Clearly described	Comments:
4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments:
4.3 Were the methods reliable? For example: -was data collected by more than one method?	☐ Reliable	Comments:

-is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?		

5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments:
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments:

5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	☐ Not sure / not reported	Comments:
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Not Sure	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Partially relevant	Comments:

5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	Comments:

6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered?	☐ Appropriately	Comments:

- Was the study approved by an ethics committee?		

As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ +	Comments:

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Starr-Keddle
	Year	2011
	Aim of study	To determine the following: 1) Are there greater levels of Phosphate and other soil nutrients at the edges of meadows? 2) Are there greater levels of Phosphate and other soil nutrients in the edges that slope downwards? 3) Is species richness and diversity greater in the edges of the meadows in comparison to the cut main parts of the meadow? 4) Are there differences in species richness, diversity and species composition between meadows that have had chemical fertiliser inputs and those which haven't?
	Study design	2
	Quality score	-
	External validity	-
Population and setting	Source population	Twelve Teesdale hay meadows with moderate species were selected/

	Eligible population	Teesdale hay meadows
	Inclusion and exclusion criteria	
	Setting	Teesdale, Co Durham.
Methods of allocation to intervention/control	Methods of allocation	In each meadow a sample was taken in the main cut part of the meadow and a sample was taken at the edge of the meadow.
	Intervention description	6 meadows had received fertiliser inputs over a number of years, whilst the other six had received no artificial fertiliser input. **All 12 had received applications of FYM.** Meadows were also selected on basis of their margins. i.e. 4 meadows had a bank sloping upwards, 4 meadows had flat edges and 4 meadows had a bank sloping downwards.
	Control/comparison description	6 fields with no artificial fertiliser input.
	Sample sizes	24 samples altogether, with 2 replicates (Fertiliser history and type of margin) Sample plots were 5m x 5m as close to centre of field as possible. A perpendicular line followed to edge where second sample was taken as close to the field boundary as possible. 10 soils sample taken per plot.
	Baseline comparisons	NA
	Study sufficiently powered	Insufficient sampling and power, only 2x replication
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Soil nutrient data Full species estimate of % cover and DAFOR for each species
	Secondary outcome measures	Species richness Shannon diversity index

		Ellenberg indicator values
	Follow-up periods	N/A one off survey
	Methods of analysis	GLM ANOVA using Inputs (fertilisers versus no fertiliser), Type of edge and Location (in field or edge of field) as environmental variables Soils and botanical data were subject to a Redundancy Analysis (RDA). The 8/12 farms in Upper Teesdale were separated out and also analysed by themselves without edge effect as a predictive variable (except for in the ordination when there was no need to balance the analysis).
Results		For meadows in Upper Teesdale Phosphate, Potassium and Magnesium levels are all significantly greater in fertilised meadows There was no significant difference between fertilised and non-fertilised meadows in terms of species richness and diversity. RDA biplot indicated that species now found more rarely in UHMs tended to be associated with the edges of the meadows and with lower phosphate levels (Alchemilla xanthochlora, Geranium sylvaticum, Cirsium heterophyllum, Sanguisorba officinalis). However there may be confounding effects of other in-field management operations, i.e. grazing/cutting etc. Irrespective of fertiliser applications the edges of fields were both more species rich and diverse than the centres ($p<0.02$)
Notes	Limitations identified by author	None
	Limitations identified by	Significant number of confounding factors (manure application, grazing intensity, initial

	review team	botanical composition) which are scarcely acknowledged in the study. Sample size is too small.
	Evidence gaps and/pr recommendations for further research	Sample should be increased and influence of net nutrients, FYM and inorganic together with grazing intensity be assessed.
	Sources of funding	North Pennines AONB

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?
Study Citation	Starr-Keddle, R.E. (2011) An investigation into soil fertility and plant species composition. North Pennines AONB Partnership – Working document 1st draft.
Study Design Category	2
Assessed by & when	CE Pinches, 7th November 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐-	Comments: 12 meadows included in study are described as having medium-high species richness. Little further description of botanical composition, management, aspect, gradient of slope etc.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Yes probably
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	Comments: Selection was entirely subjective and based on fertiliser application and edge type.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐- ☐NR	Comments: Small scale study no attempts to minimise bias, comparison group may have been compromised by ignoring nutrients added as FYM.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐-	Comments: Insufficient consideration of other management factors which explain variation.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐-	Comments: Comparison group will have been compromised by ignoring nutrients added as FYM.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Comments: Significant number of confounding factors (manure application, grazing intensity, initial botanical composition) which are scarcely acknowledged in the study. Sample size is too small.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Subjective (botanical % cover, DAFOR) and objective (soil nutrient status sampling)
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments:
3.3 Were all important outcomes assessed?	☐+	Comments: Yes

Were all important positive and negative effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NA	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: Study reports single survey
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: No, sample size too small. Replication is n=2
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: No (see above on confounding factors)
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: Broadly ok.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally		Comments: Significant number of confounding factors

valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	(manure application, grazing intensity, initial botanical composition) which are scarcely acknowledged in the study. Sample size is too small.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	Comments: Study focused on too small a sample, subjectively selected.

Evidence Table

Name of Evidence Review:	Uplands Evidence review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Starr-Keddle, R.E.
	Year	2012
	Aim of study	To answer the following questions: 1) Have Upper Teesdale upland hay meadows declined in botanical quality over time? 2) Have agri-environment schemes maintained the botanical quality of Upper Teesdale upland hay meadows? 3) Is there evidence to suggest that inorganic fertiliser use is an acceptable management option for Upper Teesdale upland hay meadows in order to maintain and sustain botanical diversity?
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Hay meadows in “upper” part of Upper Teesdale
	Eligible population	Hay meadows with available botanical data
	Inclusion and exclusion	Data from meadows included if within clearly defined geographic area upstream of High

	criteria	Force.
	Setting	Upper Teesdale, County Durham
Methods of allocation to intervention/control	Methods of allocation	Not Relevant. Data collation and analysis of botanical surveys and farm management information for Upper Teesdale Upland Hay Meadows
	Intervention description	Analysis sought to categorise meadow management by quantity of historical fertiliser inputs (5 categories) 1) Received no inorganic fertiliser only FYM (15) 2) Received FYM plus 3-6 yrs of inorganic after the baseline surveys (13) 3) Received FYM plus 7-10 yrs inorganic, mainly before baseline (16) 4) Received FYM plus 11 – 19 years inorganic fertiliser , including before baseline survey (16) 5) Received FYM plus 20 + years of inorganic fertiliser And by - presence/absence of SSSI management agreement (2 categories) - meadows which had only received FYM and meadows which had received both FYM and inorganic **prior to baseline survey** (2 categories)
	Control/comparison description	Method sought to compare early botanical surveys (the baseline surveys) with later botanical surveys (the latest survey) field by field and relate changes in composition to changes in management practices whether they had a SSSI management agreement or not, and with respect to differing histories of fertiliser addition.
	Sample sizes	98 meadows with a baseline and later survey allowing comparison (of which 43 meadows had soil data for later survey period)
	Baseline comparisons	Yes meadow data from surveys undertaken in 1970s – 1980s used as a baseline

	Study sufficiently powered	Although overall sample size is good, n = 98 the different starting points of the two groups and the extent to which sample size is influencing the significance of the results. For example, more samples would probably be needed to detect the same degree of change (and significance) in the species poor samples given that there is less potential for change over the same time period
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	No of species, No of grass and rush species, no of wildflower and sedges. Due to the need to ensure equivalence between different datasets it was necessary to reduce them down to the lowest common denominator.
	Secondary outcome measures	For each meadow in each recording period, the following were calculated : Shannon diversity index; Upland hay meadow indicator score both positive indicator score (P+) and negative indicator score (N-) together with total meadow score (TM); ADAS SS Suited Species for Nutrients and Moisture; Ellenberg Indices for fertility, moisture and pH; match to NVC community/sub community using Tablefit; Key negative indicator species; Key positive indicator species.
	Follow-up periods	Analysis spans 1970s to 2012
	Methods of analysis	Changes between the baseline and latest surveys were investigated and the differences between the management and fertiliser categories were compared, using GLM and Paired sample t tests. Detrended Correspondence Analysis and Redundancy Analysis were applied to the dataset. The analyses above were conducted on the subset of 43 fields for which there were soils data post 2002, due to the reduced sample size it was only possible to look at 3 fertilizer categories. Regression between total meadow score and Ellenberg fertility investigated and

		between Phosphorous levels ppm and Ellenberg.
Results		**General trends across 98 meadow dataset** Overall there has been a highly significant decline in the mean species no , Shannon diversity , positive indicator species score and total meadow score (all $p<0.001$) in 64% of meadows from baseline survey to latest survey. For the majority this was accompanied by a highly significant increase in SS Nutrient score and Ellenberg fertility index ($p<0.001$). There were no significant results for Ellenberg pH, Moisture of SS moisture scores, the direction of change varying significantly within the dataset. **Frequency of indicator species** There were significant declines in the frequency of 12 positive indicator species and 1 negative indicator *Bromus hordaceous* between the baseline and latest surveys. Of these, the key MG3 species, *Alchemilla vulgaris, Briza media, Geranium sylvaticum* and *Sanguisorba officinalis* were found more than 50% less frequently in the latest survey. Five species significantly increased in frequency of occurrence , Holcus lanatus, Juncus effusus, Ranunculus repens, Rhinanthus minor and Trifolium dubium. **Influence of SSSI management agreement** **Botanical variables** At **baseline** the no of wildflowers, positive indicator species, total meadow score ($p<0.001$) and no of species ($p=0.029$)were all significantly higher and the Ellenberg fertility index and SS-Nutrient scores significantly lower ($p<0.001$) in meadows with management agreements compared to those without , indicating that management agreements were well targeted to SSSI meadows of highest botanical quality and lowest

		fertility. Although this pattern was maintained in the latest survey (meadows with management agreements retaining a significantly higher no of wildflowers (p<0.012) , positive indicator scores and total meadows score (p<0.001) than those outside agreement) the declines in botanical quality between baseline and latest surveys **were more statistically significant in meadows with management agreements than those without** (typically p values of 0.001 compared to values between p =0.04 and p = 0.003). Over time the Ellenberg fertility index and SS – Nutrient scores increased significantly in both groups , although the meadows with management agreements remain significantly less fertile (p<0.001) than those without management agreements. Similarly available soil data show that phosphorous levels were significantly higher in meadows that not had a management agreement, where farmers had been allowed to add inorganic fertiliser – differences in concentrations of other macronutrients, i.e N,K and Mg were not significant. pH measurements were significantly less acid (5.8) in the meadows under management agreement than those outside (5.6)* Check with RSK as P value indicates ns. Within management agreement category there was no significant change between survey periods in SS or Ellenberg moisture indices, with SSSI meadows with MA being significantly** wetter than non MA meadows as shown by SS Moisture in both the baseline and latest survey periods. The Ellenberg moisture found SSSI meadows with MA to be significantly** wetter in the latest survey only. Whilst SSSI meadows in MA had a significantly better fit to MG3b than those outside MA, over time this fit showed a highly significant decline (p =0.001)in the SSSI meadows

		with MAs. In contrast the fit to MG8 significantly increased in both groups. Indicator species All 10 key community character species were found at significantly greater frequency in SSSI meadows with MA compared to those meadows without MA in both the baseline and latest surveys , 5 species showed significant decline in both sets of meadows over time, 2 species Succisa pratensis and Troillius europaeus only showed a significant decline in SSSI MA meadows (N.B starting frequency of these species was very low in non MA meadows), 2 species Centaurea nigra and Leontodon autumnalis showed only a significant decline** in meadows with non MA. **Historic fertiliser inputs before baseline survey** Findings were similar to the MA/non MA analysis. Botanical quality (No of wildflowers, positive indicator score and total meadow score) was significantly higher at both baseline and latest survey periods in the Teesdale meadows which had only received FYM before baseline survey and not inorganic fertiliser. Similarly measures of nutrient status (SS Nutrient score and Ellenberg fertility) were significantly lower for FYM only meadows at baseline and in the later survey period. Over time highly significant declines were detected in both groups in all measures of botanical quality with a corresponding significantly increase in measures of fertility. Soils data show that Phosphorous levels are significantly lower in FYM only meadows in contrast to meadows that received FYM and inorganic fertiliser before baseline survey. There were no significant differences in the other macro-nutrients. Findings for moisture and pH follow the same pattern as the the MA/non MA analysis, apart from pH significantly increasing through time in meadows that had FYM and inorganic at baseline. **Negative indicators**

		Over time Juncus effusus significantly increased and Bromus hordaceus significantly **decreased in FYM only** at baseline meadows, whilst Holcus lanatus and Ranunculus repens sig increased only in meadows which had recived both FYM and inorganic fertiliser at baseline. Juncus acutiflorus was found at significantly higher frequency in FYM only meadows at baseline and in the later survey period. There was no sig diff in frequency of Lolium perenne at baseline but overtime it decline sig in FYM only meadows and inc sig in FYM and inorganic meadows. **Historical fertiliser inputs** In general in the baseline survey the fertility scores were already higher where inorganic fertiliser had already been added. In the latest survey both the Ellenberg fertility index and SS Nutrient scores followed the same pattern the meadows in the first 3 categories were similar and the last two categories were significantly more fertile. Surprisingly there were few significant or predictable differences between adjacent categories, i.e 1 to 2, 2 to 3 etc. Differences in baseline composition/starting point between categories may explain these findings. More differences were apparent between meadows that had received 0-6 years inorganic fertiliser and those that received 10 years plus inorganic fertiliser. Despite the above, fertiliser category was also shown to be a significant factor in variation in the Redundancy Analysis (RDA). The biplot showed that few species were closely associated with meadows of high fertility Overall there has been a dramatic decline in meadow quality between the baseline and the latest surveys across all 5 fertiliser categories even in those receiving no inorganic fertiliser. The authors conclude that there are two possible reasons for the increase in nutrient levels in category 1 (and to 2) 1) atmospheric N deposition and 2) high nutrient content of FYM. Both factors are likely to be operating together but supply of significant P via FYM may be leading to increased fertility and associated decline in

		botanical quality. Conclusion – The study provides evidence that upland hay meadows in Upper Teesdale are declining in botanical quality whilst increasing in soil fertility using Ellenberg N as a proxy for this. Results from present study suggest meadows should receive very low inputs, no inorganic fertiliser and very low applications of FYM, that FYM application especially since N deposition alone are sufficiently high to cause a reduction in plant species richness. However note concerns that Ellenberg N may respond to grazing management and herbicide application as well as nutrients.
Notes	Limitations identified by author	Data indicate that the botanical quality of the meadows surveyed at baseline in the ESA surveys was lower compared to other datasets used. Declines between time periods for this dataset were not significant and the inclusion of this data may mask magnitude of decline in Upper Teesdale meadows included. Due to lack of data it was not possible to account for the influence of management interventions other then nutrient additions on botanical composition.

	Limitations identified by review team	Whilst the lack of available data for other management interventions was unavoidable, the discussion and conclusions reached fail to acknowledge the potential significant contributions of these factors in changes in botanical composition and Ellenberg N which as a derived measure can be affected by factors other than direct nutrient application (e.g. herbicide application, grazing intensity). As such given this is a correlative study too great an emphasis is placed on nutrient inputs being the cause effecting change.
	Evidence gaps and/pr recommendations for further research	The relative magnitude of the decline in the botanical quality (overall and indicator species) within SSSI meadows with MA and those meadows without MA needs further investigation ideally using a power analysis and/or more sophisticated multivariate techniques. These should take into account (a) the different starting points of the two groups and the extent to which sample size is influencing the significance of the results. For example, more samples would probably be needed to detect the same degree of change (and significance) in the species poor samples given that there is less potential for change over the same time period (b) the influence of the subjective groupings with respect to nutrient addition. Alternative approaches could include cluster or ordination techniques to classify fields based on floristics rather than inferred nutrient inputs (in the absence of hard evidence). This could be quantified using high nutrient species for example, or weighted Ellenberg N scores for the baseline survey. (c) The possible effects of other drivers such as changes in grazing/cutting management
	Sources of funding	Natural England

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?
Study Citation	Starr-Keddle, R.E. (2012). Upper Teesdale: changes in upland hay meadow vegetation over the past twenty to thirty years - results presented from botanical surveys. (pp. 78p., figs, tables, bibliog.). Place of publication not given: North Pennines AONB Partnership and Natural England
Study Design Category	2
Assessed by & when	CE Pinches, 22nd October 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Comments: Source population area clearly defined and general biodiversity interest described with special reference to hay meadows which are the focus of this report.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐NA	Comments: Methods seeks to collate all available data from upland hay meadows in specific geographic area.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: Yes, Method for data collation well described . Explicit inclusion criteria set focusing on data available from clear defined “upper” part of Upper Teesdale with comparable baseline and later surveys.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments: Not applicable as study uses existing datasets to investigate relationship between presence of management agreement and past fertiliser use on aspects of vegetation change.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Study focuses on historic use of inorganic fertiliser as the intervention being investigated. There was insufficient information available on other management interventions that may have an impact, e,g. timing of grazing, liming etc.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐-	Comments: The fertilizer categories represent quite crude indications of total input of NPK and there is likely to be significant variation within and across categories. This was unavoidable given lack of detailed information on rates of application of FYM and inorganic fertiliser in each year or each meadow.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐-	Comments: The role of other management interventions was acknowledged but insufficient data meant that other potentially confounding factors could not be assessed in this study (grazing intensity, shut up date, liming)
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes, setting is applicable though slightly above average altitude c 300m.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Outcome measures were derived from the original survey data available for each matched meadow pair. Frequency measures were converted to equivalent scales to enable analysis.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐NR	Comments: By its nature the study had to make best use of available data so outcome measures were not complete for all measurements, e.g soil data were available from 2002 for 43 out of the 98 fields.
3.3 Were all important outcomes assessed?	☐++	Comments: Yes, comprehensively based on available data.

Were all important positive and negative effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: The majority of outcome measures were derived, some with transformation (e.g Ellenberg values conversion from frequency to percentage cover).
3.5 Were there similar follow up times in exposure and comparison groups?	☐-	Comments: No, since the comparison is between two broad time periods 1970 – 1990 and post 2000+ hence there may exist significant differences in no of years exposed to fertiliser/FYM/other intensive management interventions in these two timeframes.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐NR	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: Sample size of 98 paired meadows satisfactory, though smaller sample sizes of no less than 13 in fertilizer categories which may be border line.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: It was not possible to account for the influence of management interventions other then nutrient additions on botanical composition. This was due to a lack of available data for the 98 meadow dataset under analyses. Whilst this was unavoidable, the discussion and conclusions reached fail to acknowledge the potential significant contributions of these factors
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: The relative magnitude of the decline in the botanical quality within SSSI meadows with MA and those meadows without MA needs further investigation ideally using a power analysis and/or more sophisticated multivariate techniques. These should take into account (a) the different starting points of the two groups and the extent to which sample size is influencing the significance of the results. For example, more samples would probably be needed to detect the same degree of change (and significance) in the species poor samples given that there is less potential for change

		over the same time period (b) the influence of the subjective groupings with respect to nutrient addition. Alternative approaches could include cluster or ordination techniques to classify fields based on floristics rather than inferred nutrient inputs (in the absence of hard evidence). This could be quantified using high nutrient species for example, or weighted Ellenberg N scores for the baseline survey. (c) The possible effects of other drivers such as changes in grazing/cutting management (as above)
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐+	Comments: Yes
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	Comments: Potentially significant confounders for key Ellenberg N derived outcome measure.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Tallowin, J.R.B.
	Year	1996
	Aim of study	To review results of series of experiments on botanically diverse meadows on the Somerset levels, examining the effects of inorganic fertiliser application on agricultural output, nutrient uptake and loss and on botanical composition N.B. Aspects of the experiment looking at restoration of former diversity following the cessation of fertiliser inputs are not discussed as they fall outside the scope of this review.
	Study design	1
	Quality score	++
	External validity	+ (Partially relevant due to study taking place on peat soils)
Population and setting	Source population	Species rich hay meadows of the NVC types MG5, MG8 and MG4.
	Eligible population	Species rich hay meadows
	Inclusion and exclusion criteria	-

	Setting	Tealham and Tadham Moor SSSIs, Somerset Levels
Methods of allocation to intervention/control	Methods of allocation	**Large scale experiment** - 3 blocks, 5 treatments randomly allocated to plots within blocks – **Small scale experiment** – 3 blocks, 19 treatments randomly allocated to plots within blocks.
	Intervention description	Large scale experiment • Fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Phosphorous and Potassium were applied in amounts to replace that removed in the hay crop. • Treatment plots were cut for hay after July 1st and the aftermath grazed by beef cattle – a compressed sward height of5.5-6.5cm was maintained during grazing period. • In 1990 plots were spilt with one half continuing to receive fertiliser inputs (N+) as previously until April 1993 whilst inputs were ceased on the other half (N-). • Experiment ran 1986 - 1993 1st Small scale experiment under cutting management only once after 1 July and again in autumn: • Fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Plus 100 or 200 kg N/ha with 0P and K replaced, 0 or 100 or 200 kg N/ha with 75kg P/ha and K replaced and 200kg N/ha, 75 kg P/ha and 200 kg K/ha. • Experiment ran from 1986 to 1989 2nd Small plot experiment set up within N0 and N200 large scale plots in 1991 - 1992 to investigate influence of cutting date and previous fertiliser treatment – NB results not presented as outside scope of this review. Cutting dates were wither late May, early-mid July, early August or in early September, then aftermath grazed

	Control/comparison description	Large scale experiment O NPK input control Small scale experiment O N with P and K replaced comparator.
	Sample sizes	Reported in primary literature refer to evaluations Kirkham Mountford & Wilkins (1996) and Mountford Lakhani & Kirkham (1993)
	Baseline comparisons	1986 first year of experiment after set up for large and small scale experiment. Small plot experiment set up within large scale plots started in 1991.
	Study sufficiently powered	Yes x 3 replication and sampling sufficient
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	Aspects of botanical composition not described in this review but fully described in primary paper evaluations Kirkham Mountford & Wilkins (1996) and Mountford Lakhani & Kirkham

significance)		(1993) Also: Hay yield Animal liveweight production and utilised metabolizable energy (UME) Soil nitrate and ammonium concentrations measured regularly between Nov 1986 and March 1990 Rates of microbial degradation of nitrate were measured between Oct 1988/89 and 1989/90
	Secondary outcome measures	Rainfall and temperature measures were taken within the experimental areas over the course of the project. Water table within each treatment plot.
	Follow-up periods	Large scale experiment ran from 1986 – 1993 Small scale experiment ran 1986 – 1990 with the influence of cutting date and previous fertiliser management measured until 1992. Small plot experiment ran 1991 – 1993.
	Methods of analysis	Details of analysis are not presented in this review paper (please refer to evaluations of individual papers on primary papers resulting from Tadham namely Kirkham Mountford & Wilkins (1996) and Mountford Lakhani & Kirkham (1993)
Results		**Effects on botanical composition reported only**

		Application of inorganic N,P and K fertiliser resulted in changes in botanical composition of the meadow communities with the extent and speed of changes reflected the amount of fertilizer applied and the management. **Large scale experiment** N inputs of 50 or more kg/ha reduced the species per m2 compared with the unfertilized controls after just three years of treatment. Addition of nitrogen fertiliser at levels as low as25 kg ha-1 yr-1 (and with other nutrients applied at very low rates only to replace that lost by hay cutting, c 9 kg P ha-1 yr-1 and c 60 kg K ha-1 yr-1) led to reduced botanical diversity and an increased abundance of competitive species, particularly the grasses Yorkshire fog *Holcus lanatus* and perennial rye grass *Lolium perenne*, after **seven** years. **Small scale experiment** Botanical diversity was lost after just one year of applying 25 (or more) kg fertilizer N/ha with P and K replaced. The most rapid and severe loss in botanical diversity occurred where 100 (or more) kg N and 75 kg P/ha were applied. Ordination studies showed that botanical change was influenced to a greater extent by application of fertilizer P than by N. Treatments that included N with both P and K (from N25 to N200) significantly reduced ($p<0.05$).Simpson's diversity index compared to the control after 1 year. **Changes in botanical composition (these are fully described in the evaluations for the primary** The botanical composition of the unfertilised control plots changed during the course of the project in both the small and large scale experiments. Greatest change, involving loss of species diversity occurred in the small scale plots not given fertilizer. A key factor in this decline in diversity appeared to be a lack of grazing in this cutting only experiment. Grazing , as Smith & Rushton 1994 has shown is essential for the maintenance of botanical diversity in many unimproved grasslands. Changes in the botanical composition of the large scale unfertilised plots during the course of the experiment may be attributable to the relatively intensive grazing management applied.

		Evidence from the controls in both the large and small scale experiments indicates that there was a decline in K availability relative to both N and P , historically inputs from FYM were applied and the decline in K may denote requirement for replenishment by this source. Fertilizer inputs particularly of P phosphorous caused increased dominance by grasses and reduction in the abundance of most of the distinctive wet hay meadow specie.
Notes	Limitations identified by author	-
	Limitations identified by review team	-
	Evidence gaps and/pr recommendations for further research	-
	Sources of funding	MAFF, DOE and English Nature

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	**What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?**

Study details	Authors	Tallowin J.R.B.
	Year	1998
	Aim of study	To collate and evaluate information relevant to the use and effects of lime application on semi-natural grasslands.
	Study design	3 (Review) Incorporates literature review and questionnaire of use of current and historical liming practice on conservation sites.
	Quality score	++
	External validity	-
Population and setting	Source population	British semi-natural grasslands
	Eligible population	As above
	Inclusion and exclusion criteria	
	Setting	

Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	
	Sample sizes	Questionnaire returns were provided for only 5 MG3 meadowss
	Baseline comparisons	N/A
	Study sufficiently powered	Sample of MG3 sites for which management data is provided is too small to be representative
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	N/A review
	Secondary outcome measures	N/A review
	Follow-up periods	N/A review
	Methods of analysis	N/A review
Results		Nutrient and liming information was presented for five MG3 grassland SSSIs from a wider survey of management practices. Of these only one SSSI had a history of lime application whilst four had a history of FYM application, with one also receiving basic slag in the past. Although based on a very small sample of MG3 SSSIs, the survey indicates that lime application is not always part of the traditional management of species rich upland hay meadows. Tallowin states 'that the generality of lime use appears to be less than that of FYM, and that the small liming effect of FYM may assist in the maintenance of this type of neutral grassland'. Furthermore 'where there is a tradition of lime use on an MG3 meadow then this tradition should continue, providing that only lime and not phosphatic slag is used'.

Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	CCW, EN, SNH, JNCC

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	**What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?**

Study details	Authors	Tallowin J.R.B.
	Year	1998
	Aim of study	To collate and evaluate information relevant to the use and effects of lime application on semi-natural grasslands.
	Study design	3 (Review) Incorporates literature review and questionnaire of use of current and historical liming practice on conservation sites.
	Quality score	++
	External validity	-
Population and setting	Source population	British semi-natural grasslands
	Eligible population	As above
	Inclusion and exclusion criteria	NA
	Setting	Not applicable.

Methods of allocation to intervention/control	Methods of allocation	NA
	Intervention description	NA
	Control/comparison description	NA
	Sample sizes	Questionnaire returns were provided for only 5 MG3 meadowss
	Baseline comparisons	N/A
	Study sufficiently powered	Sample of MG3 sites for which management data is provided is too small to be representative
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	N/A review
	Secondary outcome measures	N/A review
	Follow-up periods	N/A review
	Methods of analysis	N/A review
Results		Nutrient and liming information was presented for five MG3 grassland SSSIs from a wider survey of management practices. Of these only one SSSI had a history of lime application whilst four had a history of FYM application, with one also receiving basic slag in the past. Although based on a very small sample of MG3 SSSIs, the survey indicates that lime application is not always part of the traditional management of species rich upland hay meadows. Tallowin states 'that the generality of lime use appears to be less than that of FYM, and that the small liming effect of FYM may assist in the maintenance of this type of neutral grassland'. Furthermore 'where there is a tradition of lime use on an MG3 meadow then this tradition should continue, providing that only lime and not phosphatic slag is used'.

Notes	Limitations identified by author	The review highlights a lack of sources of information on the response of traditionally grazed semi-natural neutral grasslands to lime applications – other than the Park Grass Experiment there is no quantifiable evidence
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Author identified need for long term experimental studies on the effects of inorganic P input with and without lime applications, against equivalent inputs of P in the form of FYM on semi-natural neutral grasslands (This idea was realised in the Defra funded FYM project see Kirkham et al. in prep).
	Sources of funding	CCW, EN, SNH, JNCC

Quality Assessment Checklist: Qualitative Study v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Tallowin J.R.B. 1998. Use and Effects of Lime Application on Semi-Natural Grasslands in Britain. CCW contract science report no.FC 73-01-185.
Study Design Category	3
Assessed by & when	C.E. Pinches 20th December 2012

Section 1: Theoretical approach		
1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments: Yes, reviews collates and evaluates available evidence on the use and effects of lime application on semi-natural grasslands.
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments: Yes
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure / inadequately reported	Comments: It is not clear what approach has been applied to searching the literature for relevant evidence/information. However the number of citations referred to in the text is lengthy and indicates a comprehensive review has taken place.

Section 2: Study Design		
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure / inadequately reported	Comments: It is not clear what approach has been applied to searching the literature for relevant evidence/information. However the number of citations referred to in the text is lengthy and indicates a comprehensive review has taken place.

Section 3: Data Collection		
3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Appropriately	Comments: Not clear how references were searched for and whether this was systematic, but survey of site managers of semi-natural grassland SSSIs was carried out appropriately.

Section 4:Trustworthiness		
4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐Described	Comments: Contracted report

4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments: Yes clearly described. .
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Not Sure / not reported	Comments: Not reported

Section 5: Analyses		
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments: No explicit quantitative analysis was conducted, instead the literature is reviewed and reported.
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments: A wide diversity of literature has been used.

5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	NA	Comments: NA literature review only
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Convincing	Comments: Findings are clearly presented and well evidenced .
5.5 Are the findings relevant to the aims of the study?	☐ Relevant	Comments: Yes relevant.
5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Clear	Comments: The conclusions are clear, as is identification of areas which would benefit from research.

Section 6: Ethics		
6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered? - Was the study approved by an ethics committee?	☐ Not Sure / not reported	Comments: NA

Section 7: Overall Assessment		
As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ ++	Comments: Well conducted – despite there being no description of the method used to search the literature, the list of references evaluated and cited is comprehensive.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?

Study details	Authors	Tallowin J.R.B, Kirkham, F.W, Wilkins, R.J., Smith, R.E.N, Thomas, G. H., Mountford, J.O. & Lakhani K.H.
	Year	1994
	Aim of study	To establish; i) if there is a safe limit to the amount of fertiliser that could be applied to the species rich hay meadows of the Somerset Levels which would allow their floristic diversity to be maintained. ii) the agricultural output achievable within any such safe limit; iii) the agricultural output foregone by adhering to a 'safe' fertiliser input; N.B Significant elements of this project focused on recovery/restoration of the vegetation from fertiliser application and are not reviewed here as restoration of species rich communities is not covered by this review. Similarly the agronomic elements of the study are not evaluated falling outside the scope of the review.
	Study design	1
	Quality score	++

	External validity	+ (Partially relevant due to study taking place on peat soils)
Population and setting	Source population	Species rich hay meadows of the NVC types MG5, MG8 and MG4.
	Eligible population	Species rich hay meadows
	Inclusion and exclusion criteria	Meadows had been subjected to late hay cutting and aftermath grazing, with no inorganic fertiliser inputs. Historically the hay meadows received periodic inputs of FYM but no data were available on the frequency or rate of this input.
	Setting	Tadham Moor SSSI in the Brue Valley, Somerset Levels
Methods of allocation to intervention/control	Methods of allocation	**Large scale experiment** - 3 blocks, 5 treatments randomly allocated to plots within blocks – **Small scale experiment** – 3 blocks, 19 treatments randomly allocated to plots within blocks.
	Intervention description	**Large scale experiment** established in 1986: • Five fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Phosphorous (as Triple Phosphate) and Potassium (muriate of Potash) were applied in amounts to replace that removed in the hay crop on all plots except controls - calculated from yield and chemical analysis of hay swath samples. • In 1990 plots were spilt with one half continuing to receive fertiliser inputs (N+) as previously until April 1993 whilst inputs were ceased on the other half (N-). • Annual applications of N were split between two equal dressings, the first as soon as ground conditions allowed after mid April and the second after the removal of the hay crop. P and K were applied in mid season each year on the day following the second N application. • Treatment plots were cut for hay after July 1st and the aftermath grazed by beef cattle – a compressed sward height of5.5-6.5cm was maintained during grazing period.

		• Experiment ran 1986 - 1993 A wider range of P and K inputs were applied within a **small scale experiment** (1) under cutting management only once after 1 July and again in autumn: • Fertilizer N treatments applied annually : 0, 25, 50, 100 and 200kg/ha • Plus 100 or 200 kg N/ha with 0P and K replaced, 0 or 100 or 200 kg N/ha with 75kg P/ha and K replaced and 200kg N/ha, 75 kg P/ha and 200 kg K/ha. • Experiment ran from 1986 to 1989 A second small scale experiment (2) was set up within N0 and N200 large scale plots in 1991 - 1992 to investigate influence of cutting date and previous fertiliser treatment Cutting dates were either in late May, early-mid July, early August or in early September, then aftermath grazed

	Control/comparison description	Large scale experiment O NPK input control Small scale experiment Nill N control but with replacement P and K.
	Sample sizes	Large scale experiment - 24 x1m2 quadrats per plot (1986 – 1989) and 16 x 1m2 quadrats from (1990 – 1993) Small scale experiment - 2 x 0.5m2 quadrats per plot 1986 – 1990.
	Baseline comparisons	1986 first year of experiment after set up for large scale experiment and 1991 for small scale experiment.
	Study sufficiently powered	Yes X 3 replication – sampling sufficient
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	**Botanical** % cover of species present Mean species number per quadrat in each plot Total no of species in flower per plot was recorded in the large scale plots. **Agricultural output** Hay yield Animal liveweight production per ha

		Utilised metabolizable energy (UME) **Soil** Soil nitrate and ammonium concentrations measured regularly between Nov 1986 and March 1990 Rates of microbial degradation of nitrate were measured between Oct 1988/89 and 1989/90 Rainfall and temperature measures Water table within each treatment plot.
	Secondary outcome measures	Simpson's index of diversity
	Follow-up periods	7 years large scale experiment. 4 years – small plot experiment.
	Methods of analysis	For large scale experiment – ANOVA of each variable in each year to test the null hypothesis of equality of the experimental treatments. If the null Ho rejected then each of the 4 nitrogen application treatments was compared with the control treatment, using student's t test. Significance of linear effects of nitrogen levels was also examined. For small scale experiment - ANOVAs looking at treatment effects on different variables Relative abundance data, the number of species were used to produce dominance diversity curves. Two forms of ordination were used to relate community composition to N, P and K

		applications. CCA and DCA.
Results		**Effects on botanical composition reported only** **Large scale experiment** Species richness of the hay meadows was significantly lower than the control in the lowest fertilizer input of 25kg N per ha per annum within 6 years suggesting there is no safe amount of fertiliser N that can be applied to these meadows . Significant reduction in species number occurred within 2 years under inputs of 100 or 200kg N per ha, 3 years with inputs of 50kg N per ha and 6 years under 25kg N per ha. After 7 years of input the N200 plots supported 17 species per m2, the N100, N50 and N25 plots supported 20,20 and 24 species per m2 respectively compared with the control plots which supported 27 species per m2. There were significant reductions to the number of species in flower in the 50, 100 and 200kg N plots. A taller grass dominated sward was created in plots that received 50kg or more of N per ha. **Species changes** Of the 157 species recorded in the study area between 1986 and 1993, the abundance of 50 as significantly affected by fertilizer treatment in at least one year. Of these 13 species showed a significant increase – Agrostis stolonifera, Bromus hordaeceus, Bromus racemosus, Cerastium fontanum, Cirsium arvense, Holcus lanatus, Lolium perenne, Phleum pratense, Poa trivials, Rumex acetosa, Rumex crispus, Stellaria media and Taraxacum agg. Some low growing forbs and bryophytes disappeared locally in high N treatment plots. A large number of forb species showed a significant reduction in abundance on plots receiving fertilizer. 44 species showed a significant decrease in abundance in response to fertilizer input in at least one year, six of these were grasses, 6 were sedges and the rest were lower growing dicotyledonous species and mosses.

		The number of flowering plants of species indicative of old wet meadows declined in response to fertilizer input. Meadow thistle, Cirsium dissectum, Ragged robin, Lychnis flos cuculi, Cuckoo flower Cardamine pratensis, Lotus pedunculatus and Meadowsweet Filipendula ulmaria almost completely disappeared in plots receiving high N inputs. Prior to 1986 these were all abundant but, after 7 years of fertilizer applications they were only common on the control plots receiving no inputs. **Small scale experiment** **Responses to fertilizer N, P and K input under cutting** Botanical changes that occurred in the small scale experiment were broadly similar to those that occurred in the main plots with grasses and in particular Yorkshire Fog, Holcus lanatus becoming dominant in fertilized plots. Phosphorus was the most influential of the three elements in causing botanical change and in determining herbage production. The botanical composition of the unfertilised control plots changed during the course of the project in both the small and large scale experiments. Greatest change, involving loss of species diversity occurred in the small scale plots not given fertilizer. A key factor in this decline in diversity appeared to be a lack of grazing in this cutting only experiment. Changes in the botanical composition of the large scale unfertilised plots during the course of the experiment may be attributable to the relatively intensive grazing management applied. Evidence from the controls in both the large and small scale experiments indicates that there was a decline in K availability relative to both N and P , historically inputs from FYM were applied and the decline in K may denote requirement for replenishment by this source. Fertilizer inputs particularly of P phosphorous caused increased dominance by grasses and reduction in the abundance of most of the distinctive wet hay meadow specie. Under high fertilizer input the species rich wet hay meadow community was replaced by species poor plant community types.

Notes	Limitations identified by author	Site flooded in March 1990. Small plot experiment was not fully factorial - no treatment of P applied without K.
	Limitations identified by review team	Note this experiment was on a peat soil which are typically more deficient in plant available P compared to mineral soils.
	Evidence gaps and/pr recommendations for further research	Of those directly relevant to this project further research is needed to: Understand P availability and its effects on the recovery and maintenance of high floristic diversity Identify optimum conditions for the recruitment of seedlings of sensitive and/or rare species into these meadow communities, specifically by understanding the role of grazing.
	Sources of funding	MAFF, NCC and DOE

Name of Evidence Review: Upland Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?
Study Citation	Tallowin et al. (1994) The effects of inorganic fertilisers in flower rich hay meadows on the Somerset Levels. English Nature Research Report Number 85. Peterborough. Executive summary and summary results and conclusions of the MAFF/DOE/English Nature Tadham Moor Project 1986 -1993 (Contract F78-12-04).
Study Design Category	1
Assessed by & when	CE Pinches, 6th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	NVC communities present and soil characteristics well described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Yes, sampled area is representative of species rich wet grassland found in Somerset and in other areas of England.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	No details are provided on selection of three experimental blocks.

Section 2: method of allocation to intervention(or comparison)

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: The experiment employs a fully randomised block design, with three replicate blocks and randomly allocated 5 plots within these.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Yes all treatments well described and repeatable.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: No, none reported.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes, a traditional hay cutting and aftermath grazing regime were applied across all treatments in the large scale experiment. The details of this are well described and significant care was taken to ensure plots were grazed to the same level over the same period. Site subject to flooding which will have been variable across plots. Within the small scale experiment there was no aftermath grazing instead a two cut regime was applied.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes the meadow types present are representative of species rich wet meadows grasslands in the lowlands with a high degree of relatedness to MG3 and upland MG8 stands in the North Pennines.
2.7 Did the intervention(s) or control	☐++	Comments: Yes, the range of nutrient rates applied

comparison(s) reflect the usual UK practice(s)?		reflects range of N applied across low to high input systems.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Both - Subjective botanical assessments - % cover of each species present. % cover of litter and bare ground Height of vegetation (mean height to first flag leaf in dominant grass). Density of inflorescences for no of species of conservation interest in large scale experiment only. Mean for 24 quadrats per plot or 16 in the case of the small scale experiment. Objective - soil sampling for soil nutrients and measures of microbial community structure.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes botanical and soil measures are appropriate.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Yes, derived variables , mean plant species no across 1m2 quadrat per plot. Species richness value per 24m2 across plot or 16 m2 for small scale experiment. Species richness value as above but covering only those species in flower Simpson's index of diversity
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes.

Section 4: Analyses

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐NR	Comments:
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	Comments: No power analysis conducted but there is suitable replication of treatments and the sampling within these treatments is adequate.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Yes.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Not to a great extent in this summary report but elsewhere in the primary literature – see evaluations of Tallowin (1996), Mountford, Lakhani & Kirkham (1993); Kirkham, Mountford & Wilkins (1996).
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: Yes well designed and conducted experiment.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to	☐+	Comments: Yes, although partially relevant to MG3 hay meadows which overlie mineral soils (not peat in case of Tadham study and are known to leach considerable N).

determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?		

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: Upland Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	a) What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows ?
Study Citation	Tallowin (1996) Effects of inorganic fertilisers on flower rich hay meadows: a review using a case study on the Somerset Levels, UK. Grassland and Forage Abstracts, Vol 66 pages 147 – 152.
Study Design Category	1
Assessed by & when	CE Pinches, 6th November 2012

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	NVC communities present described in summary (full descriptions of botanical composition and soil characteristics are described in primary text (see Mountford, Lakhani & Kirkham (1993); Kirkham, Mountford & Wilkins (1996).
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Yes, sampled area is representative of species rich wet grassland found in Somerset and in other areas of England.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐NR	No details are provided on selection of three experimental blocks.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to	☐++	Comments:

management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?		The experiment employs a fully randomised block design, with three replicate blocks and randomly allocated 5 plots within these.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Yes all treatments are described in brief and repeatable – fuller descriptions are provided in Mountford, Lakhani & Kirkham (1993); Kirkham, Mountford & Wilkins (1996).
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: Yes
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐+	Comments: No, none reported.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments: Yes, a traditional hay cutting and aftermath grazing regime were applied across all treatments in the large scale experiment. The details of this are well described and significant care was taken to ensure plots were grazed to the same level over the same period. Site subject to flooding which will have been variable across plots. Within the small scale experiment there was no aftermath grazing instead a two cut regime was applied.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Yes the meadow types present are representative of species rich wet meadows grasslands in the lowlands with a high degree of relatedness to MG3 and upland MG8 stands in the North Pennines.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK	☐++	Comments: Yes, the range of nutrient rates applied reflects range of N applied across low to high input

practice(s)?		systems.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Both - Subjective botanical assessments - % cover of each species present. % cover of litter and bare ground Height of vegetation (mean height to first flag leaf in dominant grass). Density of inflorescences for no of species of conservation interest in large scale experiment only. Mean for 24 quadrats per plot or 16 in the case of the small scale experiment. Objective - soil sampling for soil nutrients and measures of microbial community structure.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes botanical and soil measures are appropriate.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐+	Comments: Yes, derived variables , mean plant species no across 1m2 quadrat per plot. Species richness value per 24m2 across plot or 16 m2 for small scale experiment. Species richness value as above but covering only those species in flower Simpson's index of diversity
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Yes.

Section 4: Analyses		
4.1 Were exposure and comparison groups		Comments: -

similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++	Comments: No power analysis conducted but there is suitable replication of treatments and the sampling within these treatments is adequate.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Yes
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: Yes.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: Not to a great extent in this summary report but elsewhere in the primary literature – see evaluations Mountford, Lakhani & Kirkham (1993); Kirkham, Mountford & Wilkins (1996).
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐++	Comments: Yes well designed and conducted experiment.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised	☐+	Comments: Yes, although partially relevant to MG3 hay meadows which overlie mineral soils (not peat in case of Tadham study and are known to leach considerable N).

across the population(s)/area(s) and nationally (i.e. habitat, species)?		

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	**What management regime maintains the diversity of the flora and fauna of the upland hay meadow Priority Habitat?**

Study details	Authors	Vickery et al.
	Year	2001
	Aim of study	To identify and outline the range of potential mechanisms by which the intensification of grassland management may impact on bird populations in Britain To review our current understanding of the mechanisms involved To highlight gaps in current knowledge about the impacts of grassland management on birds To consider ways in which grassland management could be modified to benefit grassland birds
	Study design	3 (Review)
	Quality score	++
	External validity	+
Population and setting	Source population	Generalist farmland birds (Breeding season birds) mainly waders, lapwing and songthrush or passerines. In winter foraging waders and passerines.
	Eligible population	British bird populations

	Inclusion and exclusion criteria	Impacts of drainage (and its' impact on invertebrate prey abundance), predation and roosting and feeding sites for wildfowl together with re-seeding were excluded from the review. The impact of pesticides was also excluded, although the impact of avermectins was considered under grazing management. Wildfowl were excluded.
	Setting	UK
Methods of allocation to intervention/control	Methods of allocation	Review considered broad components of intensification; fertiliser use; stocking practices and cutting regimes
	Intervention description	NA review
	Control/comparison description	NA review
	Sample sizes	NA
	Baseline comparisons	NA
	Study sufficiently powered	NA
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	NA
	Secondary outcome measures	NA
	Follow-up periods	NA
	Methods of analysis	NA
Results		Note: information pulled out of the reviews results is limited to evidence not identified in the primary references revealed by search. Other information presented in the paper has been used in the context sections but not in support of evidence statements, as the primary literature

		on which this review is based has been referred to directly. **Effect of nutrient form** Vickery *et al.* (2001) [3++] state that moderate use of FYM may benefit grassland birds by increasing the abundance of soil-dwelling invertebrates, or their accessibility by bringing them closer to the surface. They report that winter field use by lapwings, starlings, redwing and fieldfare is positively associated with frequent addition of FYM on permanent grassland but that benefits decrease under high applications and would be expected to decrease if the livestock have been recently dosed with broad-spectrum avermectin wormers.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Three areas are identified as being of particular need of further research to improve the evidence base, they are: (i) the inter-action between changes in food abundance, due to changes in fertilizer inputs, and food accessibility, due to changes in sward structure; (ii) the interaction between predation rates and management-related changes in habitat and (iii) the impact of alternative anti-helminithic treatments for livestock on invertebrates and birds.
	Sources of funding	MAFF and JNCC

Quality Assessment Checklist: Qualitative Study v2.0

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay meadows

Review Question	What types, rates of application and timing/periodicity of nutrient and lime applications maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Vickery, J. R. Tallowin, R. E., Feber, E. J., Asteraki, P. W., Atkinson, R. J., Fuller & V. K. Brown (2001) The Management of Lowland Neutral Grasslands in Britain: Effects of Agricultural Practices on Birds and Their Food Resources Journal of Applied Ecology, Vol. 38, No. 3 pp. 647-664
Study Design Category	3
Assessed by & when	C.E. Pinches 20th December 2012

Section 1: Theoretical approach		
1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question? C	☐ Appropriate	Comments: Yes, reviews available evidence on on the various mechanisms by which the intensification of grassland management may impact on bird populations in Britain focusing on lowland neutral grassland. It seeks to examine the link between grassland management and its use by generalist farmland birds for nesting and, in particular, for foraging, rather than the impact of management on a particular bird species.
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments: Underpinning assumptions clearly set out as are what is and isn't included. The review focuses on the impact of three broad components of intensification were considered: fertilizer use, stocking practices and cutting regimes.
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure / inadequately reported	Comments: It is not clear what approach has been applied to searching the literature for relevant evidence/information. However the number of citations referred to in the text is lengthy and indicates a comprehensive review has taken place.

Section 2: Study Design		
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure / inadequately reported	Comments: It is not clear what approach has been applied to searching the literature for relevant evidence/information. However the number of citations referred to in the text is lengthy and indicates a comprehensive review has taken place.

Section 3: Data Collection		
3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Appropriately ☐ Not Sure / inadequately reported	Comments: Not clear how references were searched for and whether this was systematic.

Section 4:Trustworthiness		
4.1 Is the role of researcher clearly described? For example: -has the relationship between the researchers and intervention group been adequately considered?	☐ Clear	Comments: Though institution/organisation to which authors/researchers belong is clear under the author names and the funders MAFF and JNCC are acknowledged.

4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments: Yes context of declining farmland bird populations and changes to grassland management very clear described.
4.3 Were the methods reliable? For example: -was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?	☐ Not Sure / not reported	Comments

Section 5: Analyses		
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Sure / not reported	Comments: No explicit quantitative analysis was conducted, instead the literature is reviewed and reported. A key finding of the study was that few interactions between grassland management and changes in faunal populations have been quantified.
5.2 Is the data 'rich'? For example: -how well are the contexts of the data described? -has the diversity of perspective and content been explored? -are responses compared and contrasted?	☐ Rich	Comments: A wide diversity of literature has been used.

5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	NA	Comments: NA literature review only
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Convincing	Comments: Findings are clearly presented and well evidenced .
5.5 Are the findings relevant to the aims of the study?	☐ Relevant	Comments: Yes relevant.
5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Clear	Comments: The conclusions are clear and any areas of speculation are acknowledged as are further areas which would benefit from research.

Section 6: Ethics		
6.1 How clear and coherent is the reporting of ethics? For example: -have ethical issues been taken into consideration? -Are they adequately considered? -Have the consequences of the research been considered? - Was the study approved by an ethics committee?	☐ Not Sure / not reported	Comments: NA

Section 7: Overall Assessment		
As far as can be ascertained from the paper, how well was the study conducted? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ ++	Comments: Well conducted – despite there being no description of the method used to search the literature, the list of references evaluated and cited is comprehensive.

Quality Assessment Checklist: Quantitative Study Observational / Correlation v2.0

Name of Evidence Review: __Upland______________________

Name of Review Sub-topic (if any): ___Upland hay Meadow______

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Wilson, R. (1991) Yellow wagtails in Littondale and Arkkengarthdale. English Nature Report;. North East region.
Study Design Category	3
Assessed by & when	CE Pinches 20th December 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐NR	Not in this study but description would not have been necessary as this was a commissioned survey and report designed to investigate cutting date and fledging survival of yellow wagtails.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Yes, two dales sampled – providing a degree of contrast in terms of altitude and climated.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Yes.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Survey approach rather than comparison study. All fields selected for presence of breeding birds.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	Only effect of cutting date investigated.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NA	Not applicable survey
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Effect of cold spring delaying arrival and nesting of birds was discussed
2.5 Is the setting applicable to the UK?	☐+	Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measures subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Yes
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	As far as they could be although losses prior to cutting due to predation etc could not be ruled out based on 4 visits.

3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Yes number of yellow wagtails successfully raising fledglings.
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Yes

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NA	Descriptive study
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐NA	Descriptive study
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐NA	Descriptive study
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NA	
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)?	☐+	Yes

Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	Yes broadly, although cold atypical nature of spring in season of survey makes them less applicable, hence – score.

Evidence Table

Name of Evidence Review:	Uplands Evidence Review
Name of Review Sub-topic (if any):	Hay Meadows
Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?

Study details	Authors	Wilson, R
	Year	1991
	Aim of study	To more accurately assess the timing of fledging in relation to the cutting of silage/hay crop within two Dales in the Yorkshire Dales National Park, as a follow up to wider scale survey in the previous year (1990).
	Study design	3
	Quality score	+
	External validity	-
Population and setting	Source population	Nesting pairs of yellow wagtails in Littondale and Arkengathdale
	Eligible population	Yellow wagtails in Pennine Dales and other upland areas of Northern Britain.
	Inclusion and exclusion criteria	A complete survey, on foot of delineated areas known to support breeding yellow wagtails in the past was undertaken between the 23rd Mary and 6th June, This determined where male Yellow Wagtails had established or were attempting to establish territories. Thereafter only those sites where the birds had been recorded were monitored, although occasional spot-

		checks of additional sites recorded positive in the 1990 survey were made.
	Setting	Littondale (lies outside original Pennine Dales ESA) and Arkengathdale (lies within Pennine Dales ESA).
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	NA
	Control/comparison description	
	Sample sizes	Each dale was visited on four occasions after the initial survey to establish breeding territories.
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Behaviour of adult birds where indicative of a particular stage in the breeding cycle was recorded. Nesting habitat as defined by five categories was recorded. 1) Flower rich fields suggesting use as hay meadow 2)Flower poor grasslands of long grass, perhaps for silage 3) Short grass 4) Rush pasture and 5) NA bird flying over of habitat indeterminate. Numbers of Yellow wagtail present were recorded.
	Secondary outcome measures	
	Follow-up periods	Within one season – 4 visits.
	Methods of analysis	None – results simply presented.
Results		A nest failure rate for attributable to early cutting of up to 33% (13 sites) was reported. A combination of a cutting date restricted to the 15th July and a more normal year

		weather wise would extend the available breeding season and allow ample time for completion of first broods to fledging and perhaps enable some of the earlier birds to get two broods away.
Notes	Limitations identified by author	Author notes limitations of using current appearance of a field , which largely reflects past practice which may not well be repeated in any one year. There was a very cold and late spring in 1991 - the effect of the cold weather may have contributed to the losses as the birds were a full month late in nesting but hay/silage cutting was delayed by approximately 2 weeks/
	Limitations identified by review team	-
	Evidence gaps and/pr recommendations for further research	-
	Sources of funding	English Nature

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Upland Hay Meadows

Review Question	c) What spring grazing levels, timing of shut up/closure for hay and cutting dates maintain the floristic diversity and breeding bird populations of upland hay meadows?
Study Citation	Wilson, R. (1991) Yellow wagtails in Littondale and Arkkengarthdale. English Nature Report;. North East region.
Study Design Category	3
Assessed by & when	CE Pinches, 12th December 2012

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐+	
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NR	Comments:
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments:
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐NR	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐NR	Comments:
2.5 Is the setting applicable to the UK?	☐+	Comments:

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments:
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐+	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative	☐+	Comments:

effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments:
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: yes but no quantitative statistical analysis possible due to paucity of earlier data.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐NR	Comments: NA
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐NR	Comments: NA
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐+	Comments: Yes

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐-	Comments: Study is a snapshot of breeding in one season which makes extrapolation difficult.

Study Details	Population and setting	Methods of allocation to intervention / control	Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Results	Notes
Authors: Robert Wolton **Year:** 2000 **Aim of study:** To investigate the effectiveness of different cutting regimes as a method of controlling *Juncus effusus*	**Source population:** MG10a, *Holcus lanatus-Juncus effusus* rush pasture on heavy acidic clay, waterlogged for 6 months of the year **Setting:** Locks Park Farm, Hatherleigh, North Devon	**Methods of allocation:** 8 blocks were set up in a field, and 8 treatments were applied to 10 rush tussocks in each block. The methods do not state whether the allocation was random **Intervention description:** 1. Date of first cut (May, July, August, September). 2. Cutting height (flush with the ground or 8 cm above ground). 3. Fortnightly or monthly cuts	**Primary outcome measures:** % of shoots growing a month after cutting **Secondary outcome measures:**	Cutting flush with the ground is more effective than cutting at a height of 8 cm. If only a single cut is possible, then cutting after midsummer is more effective than before midsummer, but this is not the case if more cuts are made. Cutting at monthly intervals in some instances appears more effective than cutting fortnightly.	**Limitations identified by author:** The trials were carried out over only one year, so the results should be regarded as provisional **Limitations identified by review team:** 1. No statistics other than the use of 95% confidence intervals. The CIs don't clearly show a significant effect, so without a statistical test the likelihood of effects

Study design: Randomised control trial **Quality Score: -** **External validity: -**		**Control / comparison description:** No cutting **Sample sizes:** 8 replicates **Baseline comparisons:** NA - 100% shoots before cutting **Study sufficiently powered:** No power analysis. Confident intervals indicate that some difference in growth following treatments are significant	**Follow-up periods:** Growth of shoots analysed 1 month after cutting. Trials were carried out from May to November **Methods of analysis:** No analysis mentioned in the methods. Results give diagrams with confidence intervals		of the treatments cannot be ascertained. 2. Each of the replicates replies on a sample of only 10 rush tussocks, and the number of shoots per tussock before cutting began differed widely. **Evidence gaps and recommendations for further research:** **Sources of funding:** Author an English Nature employee

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Hay Meadows

Review Question	b - effectiveness of cutting to control rushes

Study Citation	Wolton, R. (2000). The control of soft rush *Juncus effusus* by cutting. Journal of Practical Ecology and Conservation 4 (1) 18 - 26
Study Design Category	1 Randomised control trial
Assessed by & when	Kate Fagan 30-11-12

1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐++	
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?**	☐+	Yes, but only one small site investigated

e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?		
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	Selection method not described.

2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐NR	Comments: Not reported whether or not selection of tussocks was random
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments:
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias?	☐+	One growth season is not sufficient time to indicate lasting effect, but it is a good initial indication

Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)		
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	A study of just one site is not fully representative, but the site is English

2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments:

3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Outcome measures objective, but huge inequality in tussocks initially
3.2 Were all outcome measurements complete?	☐++	Comments:

Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Just % growth of shoots
3.4 Were outcomes relevant?		Comments:

If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐NA	
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Longer would have been better

4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐-	**This isn't really reported, but given the reported overall large difference in tussocks at baseline, it is unlikely that there was much equality**
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐+	**There probably was sufficient replication, but no statistical tests were applied**

4.3 Were the estimates of effect size given or calculable?	☐+	**Diagrams of 95% confidence intervals**
4.4 Were the analytical methods appropriate?	☐-	**No analysis other than the calculation of the confidence intervals**

Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?		
4.5 Was the precision of the intervention [treatment?] effects given or calculable? Were they meaningful?	☐NA	**Comments**:
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	Confidence in the results can't rely on confidence intervals when they don't show clear differences - a statistical test is necessary
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)?		Comments:

Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐ -	