Evidence Table

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

Study details	Authors	Adamson, J & Kahl, J
	Year	2003
	Aim of study	To review a series of long term grazing exclosure plots, collate data and conduct repeats of vegetation monitoring at 10 sites
	Study design	2 Non-randomised controlled trials or controlled before and after study
	Quality score	+
	External validity	++
Population and setting	Source population	Extensive blanket bog and grassland swards in the north Pennines
	Eligible population	Ten exclosures erected on a range of upland vegetation types
	Inclusion and exclusion criteria	Grazed control plots located adjacent and on similar vegetation, altitude, aspect etc
	Setting	North Pennines, English Uplands
Methods of allocation	Methods of allocation	Not described but not random

to intervention/control	Intervention description	Ten stock-proof exclosures erected for 44-46 years
	Control/comparison description	Control plots identified adjacent to the exclosures. Similar in size, initial vegetation, altitude, aspect etc
	Sample sizes	Ten exclosures varying in size. Vegetation sampling covers almost 300,000 records mostly species records from points within quadrats.
	Baseline comparisons	Full vegetation recording at erection of exclosures and subsequent re-recording at different periods. These range between 1 year and 46 years
	Study sufficiently powered	This review paper is well powered Relatively small number of plot samples (10) but comprehensive and systematic measurement of vegetation within these study sites gives many thousands of vegetation measurements. The study is also well powered relating to timescales- these are long-term experiments run over five decades.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Change in vegetation following cessation of sheep grazing. Change in vegetation distribution, composition, structure. Recording at intervals gives change over time.
	Secondary outcome measures	
	Follow-up periods	Sampling at periods between 1 and 46 years. This study provided 2003 recording for all the study plots.
	Methods of analysis	% change in species at point quadrat samples. Significance of change statistically tested.

Results		The 10 long term plots covered a range of upland vegetation types. This included Agrostis-Fescue grassland, Nardus grassland, Juncus squarrosus grassland, calcareous flush, Eriophorum dominated blanket bog and Calluna-Eriophorum dominated bog. The low altitude (550-600m)deep peat sites showed a small response to exclusion of grazing. These sites however had low intensity grazing and the least extreme climate. At one site there was an increase in Calluna and a decrease in Eriophorum. The low altitude grasslands (550-640m) were very heterogeneous and included Nardus and Juncus squarrosus vegetation, as well as calcareous flushes. All sites showed decline in Festuca and Nardus and there are dramatic Juncus squarrosus declines at one site. The most visible response is an increase in forbs at the expense of monocotelydons. The higher altitude blanket bogs (690m) show a dramatic response to cessation of grazing. These plots and their controls are hard grazed an impoverished making vegetation response marked. Calluna established itself within the plots despite being well above the prevailing altitude for the species. Bare ground reduces and Empetrum, Rubus chamaemorus and Narthecium all increased. The high altitude grasslands (690-830m) are hard grazed and are exposed. Vegetation response to grazing removal is again dramatic. All sites show Deschampsia flexuosa increase (possibly responding to increased N from N deposition as well as grazing removal) Carex bigelowii increased significantly. This is at the expense of Festuca ovina, Nardus and J. squarrosus. There was a clear increase in biomass at all sites and a reduction of moss cover (prob related to vegetation depth and shade). There were increases in lichen cover attributed to a reduction in trampling. The Authors describe the “benefits to diversity and biomass” following vegetation removal. The dynamics of vegetation following change in management are not summarised but it is clear from this review paper that stability may take decades to establish and that

		changes to vegetation cover may be slow and small. For example on one blanket bog site, Calluna had established but was at low frequency after 31 years of grazing removal.
Notes	Limitations identified by author	Exclosures restricted sheep grazing but there was noted impact by voles and red grouse.
	Limitations identified by review team	No information was presented on the density of grazing livestock in the controls and prior to the exclosures being erected. It is inferred that this was a heavy grazing pressure, partly due to the dramatic response of vegetation but this is not recorded. The paper summarises grazing levels but these relate to a single point in time and it is known that these have varied significantly. Thus these studies limit themselves to the impacts of grazing or no grazing. It might be reasonable to infer that proportional changes in grazing levels would deliver some of the changes seen but this is again not scientifically proven.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	CEH, ITE, English Nature and predecessor bodies.

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

Name of Evidence Review: Uplands Evidence Review

Name of Review Sub-topic (if any): Moorland Grazing

Review Question	
Study Citation	Adamson & kahl
Study Design Category	Non-randomised controlled trials/controlled before and after studies 2
Assessed by & when	Simon Webb 9/12/12

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 source population is a range of upland vegetation types in the English uplands. This included Agrostis-Fescue grassland, Nardus grassland, Juncus squarrosus grassland, calcareous flush, Eriophorum dominated blanket bog and Calluna-Eriophorum dominated bog. in the English uplands. The paper very briefly describes the habitat but provides detailed description of each of 10 study sites
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: The trial plots are representative of Agrostis-Fescue grassland, Nardus grassland, Juncus squarrosus grassland, calcareous flush, Eriophorum dominated blanket bog and Calluna-Eriophorum dominated bog. in the English uplands. The study sites are representative of the condition of these habitats in England. Other Upland/Pennine habitats were not discussed including heathland communities.
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: This is a review paper looking at sites previously studies by other authors and repeating long-term measurements. There is some inherent selection bias here- the authors could only select from sites where baseline/historic data had been gathered. Expressed a different way this paper repeats any selection bias seen in earlier experiments (if any) The greater number and larger size of study sites provides reasonable representation of the source population.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: This is the same as 1.3 above. T There is some inherent selection bias here- the authors could only select from sites where baseline/historic data had been gathered. Expressed a different way this paper repeats any selection bias seen in earlier experiments (if any) The greater number and larger size of study sites provides reasonable representation of the source population. This review paper brings together data from numerous previous studies. The greater number of study sites and greater area would reduce bias. =
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Yes. Change in vegetation distribution, composition, structure.
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: Yes Exclosures restricted sheep grazing. Therefore impacts limited to exclusion of domestic livestock.
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: These were not given consideration in the paper. However the simple design and appropriate control plots does control much of any potential bias.
2.5 Is the setting applicable to the UK?	☐++	Comments: Study completed in UK Good representivity of the Pennine habitats.

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. Sampling intensive and repeatable. Objective-measurement rather than subject observation. DM – pin hit measurements were not always made at baseline
3.2 Were all outcome measurements		Comments:

complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Yes but.... Yes sufficient measures were completed to identify response in vegetation. Long timescale of experiment adds value and reduces short-term response bias.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Yes- as defined by the scope of the experiment
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐NR	Comments: Direct measurement of the variables were taken rather than surrogate measures
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Yes. These are long term experiments running for upto 50 years.

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 power calculation presented This can be considered on a number of different levels. The study is well powered when the number of measurements is considered. Here there are many thousands of vegetation data collected. The study is also well powered relating to timescales- these are long-term experiments run over 5 decades. The study is also well powered as a number of exclosures is considered. This review paper is well powered
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: Multiple variables were not considered in the statistical analysis but they are identified and discussed in the paper
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐++	Comments: Yes

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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
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: Main bias was in plot selection where little evidence given so show objectivity. As this is a review paper any selection bias in earlier work is repeated in this review. There did not appear to be any significant flaws in 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: The findings are applicable to the grazing management of the Pennines and other grassland/bog habitats. This is a review paper gathering data from other studies and completing a 2003 repeat on all the work It looks at a good range of habitats and runs for more than 5 decades. Its results are compelling.

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

Name of Evidence Review: _____Uplands_____

Name of Review Sub-topic (if any): _____Effects of Livestock grazing on mooralnd_____

Review Question	Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?
Study Citation	Albon, S.D., Brewer, M.J., O'Brian, S., Nolan, A.J. & Cope, D.(2007) Journal of Applied Ecology 44. Quantifying the grazing impacts associated with different herbivores on rangelands
Study Design Category	
Assessed by & when	D Martin 6th 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.	☐++ ☐+ ☐- XNR ☐NA	Comments: This is a wide study, the source population being the open hill areas of the Highlands of Scotland
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?	X ++ ☐+ ☐- ☐NR ☐NA	Comments: 11 large Deer Management Areas ranging from 148 km^2 to 1600 km^2. Each contain the seven upland/ montane habitats assessed in varying proportions.
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?	☐++ ☐+ X- ☐NR ☐NA	Comments: Yes. In 3 areas the sample covered whole area. Elsewhere a stratified random approach was used to ensure coverage of habitats and land management units, giving a sample coverage of between 12% and 21%. All habitats represented in each area.

2

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ X+ ☐- ☐NR ☐NA	Comments: Not clear how the 11 areas were chosen, and where they sit in the range of exposure (grazing pressure/ impact). Within areas sampling of exposure was minimised by either surveying whole area or stratified random sampling.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	X++ ☐+ ☐- ☐NR ☐NA	Comments: Wide range of explanatory variables – presence of different wild and domestic grazers and their interactions
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?	☐++ ☐+ ☐- XNR ☐NA	Comments: Treatments are not imposed, so issues of contamination not really relevant. Presence of the different grazing species may be subject to error/ miss-identification.
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?	☐++ X+ ☐- ☐NR ☐NA	Comments: As this is survey of prevailing grazing regimes, there were differences in timing and duration of livestock grazing, particularly sheep, numbers were adjusted to year equivalents and averaged across polygons where presence was recorded. Deer counts similarly averaged. Recording of grazing species often relies on signs of presence – may be miss-identified or under-recorded.
2.5 Is the setting applicable to the UK?	☐++ X+ ☐- ☐NR ☐NA	Comments: Yes – based in Scottish Highlands. Many of the vegetation types (except montane to a significant extent) found in English uplands, although will be differences in some of the key communities and their composition.

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?	☐++ X + ☐- ☐NR ☐NA	Comments: Grazing impacts assessed on a five-point scale through measurement of a range of field indicators, some of which have a degree of subjectivity or estimation. It is however a standard method (MacDonald et al 1998) that has been extensively field-trialled.
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?	X ++ ☐+ ☐- ☐NR ☐NA	Comments: Yes, based on the sampling strategy adopted.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ X + ☐- ☐NR ☐NA	Comments: Surrogate measures of grazing impact and grazing animal occupancy. Herbivore occupancy assessed on recorded presence, but no estimate of degree of use by each herbivore.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ X + ☐- ☐NR ☐NA	Comments: Yes, although question over whether presence of the relevant herbivores will have been consistently picked up.
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- XNR ☐NA	Comments:

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- XNR ☐NA	Comments: A survey rather than treatment approach, so assessing the prevailing conditions at a point in time. Different areas surveyed in different years (1997 -2003) so may be seasonal affects of the impacts of graziers in relation to productivity etc.

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?	☐++ ☐+ ☐- XNR ☐NA	Comments: No power analysis presented, but sample size, in terms of number of polygons surveyed is large. Sufficiently powered to detect grazing effects of the most common herbivore-habitat interactions, but not for some of the less common ones.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	X ++ ☐+ ☐- ☐NR ☐NA	Comments: Wide range of explanatory variables – presence of different wild and domestic grazers and their interactions. The key explanatory variable, in terms of likely main grazing species and major habitat interactions, were considered. Baysian regression analysis used to identify variables.
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?	X ++ ☐+ ☐- ☐NR ☐NA	Comments: Modelling approach to identify important explanatory variables. Model included a range of environmental and ecological co-variables to increase confidence in the estimates of herbivore effects. Regression of impact score vs deer and local deer density for two areas, and impact score vs regional deer density for four of the most extensive habitats.
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?	☐ ++ X + ☐- ☐NR ☐NA	Comments: The output is the estimated change in probability of observing an impact class of 'moderate' or greater with the recorded presence of a herbivore. Predicted impacts and credible range is presented, with indication of significance (range excludes 0).

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	☐++ X + ☐-	Comments: Large-scale study with large number of observations (median 1067 habitat polygons per area), usually giving enough polygons with herbivores species not recorded to allow herbivore effects to be estimated. The model predicted higher impacts in the absence (not recorded) of deer in some cases, which is not fully explained.
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)?	☐ ++ X + ☐-	Comments: wide scale study across a number of extensive upland areas with main upland habitats well represented. Methods readily applicable in other areas. Sites all Scottish Highlands, so not fully representative of the UK resource. Impact of sheep not found to be density dependant, so model has greater predictive power of deer, less common in England.

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Moorland Grazing
Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

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:	Source population: Eligible Population: Inclusion & exclusion criteria: Setting:	Methods of allocation: Intervention description: Control / comparison description: Sample sizes: 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/pr recommendations for further research: Sources of funding:
Authors: Albon, S.D.,	Source population:	Methods of allocation: Treatments not allocated,	Primary outcome measures: Recorded	Recorded presence of sheep	Limitations identified by author: Possible

Brewer, M.J., O'Brian, S., Nolan, A.J. & Cope, D. 2007 Aim of Study: To quantify the grazing and trampling impacts associated with six different herbivore species on semi-natural habitats; to explore whether it is possible to detect differenced in impact of sheep and red deer; to investigate the	The extent of unenclosed hill land in the Scottish Highlands Eligible Population: 11 large Deer Management Areas ranging from 148 km^2 to 1600 km^2. Each contain the seven upland/ montane habitats assessed in varying proportions. Inclusion & exclusion criteria: None specified Setting: Deer Management	measuring the prevailing grazing levels and patterns Intervention description: Intervention is grazing by wild herbivores (red deer, rabbits, mountain hare, red grouse) and livestock (sheep, cattle). Control / comparison description: Not a control study. Comparisons made across a number of areas. Sample sizes: Large sample: 700-3400 vegetation polygons per area, across 11 areas Baseline comparisons: Not a baseline and resurvey, but a census of a number of upland areas Study sufficiently powered No power analysis presented, but sample size, in terms of number of polygons surveyed is large. Sufficiently powered to detect grazing effects of the most common herbivore-	occupancy of different herbivores based on signs of presence; herbivore grazing and trampling impact through field indicators of structure and biomass removal. Secondary outcome measures: Follow-up periods: A survey approach rather than imposing treatments. The 11 sites were surveyed over a period 1997-2003, with 1-3 sites surveyed each year. Methods of analysis: Modelling approach to identify important explanatory variables. Model included a range of environmental and ecological co-variables to increase confidence in the	associated with higher grazing and trampling impacts than other mammalian herbivores assessed. Sheep also associated with highest impact averaged across habitats in 7 of 11 areas, and increased the probability of recording a 'moderate' or greater impact on most habitats. Presence of cattle next most likely to be associate with increased impacts, but presence more localised. Wild herbivores had comparatively little impact at the DMG (area) scale. Impact of sheep greater than deer	issues with identifying presence of different grazing species conclusively; only presence recorded, not estimates of relative density; some species- habitat interactions could not be included in the model for some areas due to low number of records. The negative association with deer in some areas may be a limitation of the model. Limitations identified by review team: Evidence gaps and/or recommendations for further research: Response of red deer in terms of spatial grazing and impact a sheep numbers fall. The evidence from individual studies

relationship between impacts and stocking rates of sheep and deer. Quality Score: 2+ External validity: 2+	Group areas covering sporting estates and sheep grazing in upland Scotland	habitat interactions, but not for some of the less common ones.	estimates of herbivore effects. Regression of impact score vs deer and local deer density for two areas, and impact score vs regional deer density for four of the most extensive habitats	in almost all area-habitat combinations. Deer impact increased with density in the two areas measured, but not sheep, which had high impact at low density. May be related to greater aggregation and smaller range size. In some area; habitat patches presence of deer associated with lower impact than when no deer recorded- could be that deer are attracted to more productive areas, which appear lightly grazed.	around the world indicates that the magnitude and direction of the effects of different herbivores varies over spatial and temporal scales. Models describing the interactions of herbivores and plant diversity need testing at a range of spatial scales. Sources of funding: SEERAD funding of synthesis and paper production, based on methods and surveys developed under SNH and Deer Management Group funding.

Evidence Table

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. Effect of grazing on delivery of biodiversity h. effects of absence/abandonment

Study details	Authors	Amar et al
	Year	2011
	Aim of study	To test whether reductions in sheep numbers have led to an increase in hen harrier prey or preferred foraging habitat, and whether breeding output correlates with sheep stocking numbers or variations in weather conditions (rainfall and temperature)
	Study design	Quantitative observational
	Quality score	+
	External validity	+
Population and setting	Source population	Orkeny, Scotland. Grid references given Moorland habitat/rough grassland – no further description provided
	Eligible population	Not reported
	Inclusion and exclusion criteria	Sites with long-term monitoring of hen harrier numbers

	Setting	Orkney, Scotland (West Mainland)
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	Varying sheep stocking densities
	Control/comparison description	N/A - Single study site used
	Sample sizes	Sample size: 2 line transects within 18 1 km squares 25x25cm quadrats every 40cm on transects (50 quadrats per square) for prey & grassland surveys
	Baseline comparisons	N/A
	Study sufficiently powered	Not reported
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Hen harrier numbers – sheep numbers
	Secondary outcome measures	prey/preferred foraging habitat/weather variables (rainfall/temperature)
	Follow-up periods	1975-2008 for hen harrier & sheep numbers 1999/2000 & 2008 for vole/lagomorphs and rough grassland surveys
	Methods of analysis	Linear regression for sheep numbers & young fledged Generalised Linear Mixed Model for changes in abundance of prey and number of rough grassland dominated quadrats

		General Linear Model for weather variables and sheep abundance
Results		An increase in the number of rough grassland quadrats (p=0.04) corresponded with an increase in vole signs (p=0.01) but no difference in lagomorphs signs (p=0.44) or meadow pipit (p=0.26) No relationship was found between total young fledged and spring or summer temperature or summer rainfall. A significant negative association was found spring rainfall. A highly significant relationship with sheep abundance was found. These 2 variables accounted for nearly 40% of variation between years
Notes	Limitations identified by author	Agricultural census data does not allow sheep numbers in an individual habitat to be determined, therefore whether data reflects study site is uncertain – anecdotal evidence suggests numbers were never high in moorland nesting areas Caution is needed when comparing two points on a time series as change between points may be due to variation around a long term trend rather than the trend itself The study may be less applicable where deer grazing is present
	Limitations identified by review team	Influence of grazing levels on vegetation type and structure (& therefore vole numbers) will depend on original vegetation, this is not reported – decreases in sheep numbers may favour dwarf shrub where it is already present, and conversely may disadvantage vole/meadow pipit populations
	Evidence gaps and/pr recommendations for further research	Study refers to single site, and is not replicated
	Sources of funding	RSPB and SNH

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

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): _____Grazing_____

Review Question	a. Effect of grazing on delivery of biodiversity h. effects of absence/abandonment
Study Citation	Amar et al (2011)
Study Design Category	Quantitative observational
Assessed by & when	Susanna Phillips 06/11/12

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.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Orkeny, Scotland. Grid references given Moorland habitat/rough grassland – no further description provided
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 ☐NA	Comments:
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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Surveys of Orkney voles, lagomorphs and meadow pipits and rough grassland – line transects within 1 km squares (selected non-randomly, locations remained similar between years to meet requirements of another study). Edge of square for transect selected randomly 25x25cm quadrats every 40cm on transects

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Single study site used
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Hen harrier numbers – prey/preferred foraging habitat/weather variables (rainfall/temperature). Previous work shows these variables affect hen harrier success due to the effect on prey and nestling mortality
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 ☐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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Supplementary feeding experiment in 1999/2000 – data for these years excluded in analysis of productivity Discrepancy between survey dates for meadow pipits but no correlation found between abundance and date Additional data used to determine whether changes in vole numbers were real or reflected larger scale temporal fluctuations
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: UK based study

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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Hen harrier numbers systematically monitored (total number of young recorded annually – to provide information on number of variables including numbers of breeding females, breeding success rate and brood size at fledgling) – details of methods not provided. June agricultural census data for sheep numbers – accuracy for specific site is uncertain Weather data from Kirkwall weather station (20km from study site) Vole and lagomorphs abundance measured as a proxy of presence or absence of droppings Meadow pipits – standard passerine transects Rough grassland – subjective measure of build up of dead vegetation forming litter mat
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 ☐NA	Comments: Weather data missing for some seasons, and were estimated using predictive linear regression from Lerwick weather station (100 miles north)
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data appropriate to meet objectives of study
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Vole and lagomorphs abundance measured as a proxy of presence or absence of droppings – appropriate measure for relative change between years June agricultural census data for sheep numbers – accuracy for specific site is uncertain
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+	Comments:

	☐- ☐NR ☐NA	
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 1975-2008 for hen harrier & sheep numbers 1999/2000 & 2008 for vole/lagomorphs and rough grassland surveys

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 ☐NA	Comments: Power calculation not present Sample size: 2 line transects within 18 1 km squares 25x25cm quadrats every 40cm on transects (50 quadrats per square) for prey & grassland surveys
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: sheep numbers/prey/preferred foraging habitat/weather variables (rainfall/temperature)
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 ☐NA	Comments: Linear regression for sheep numbers & young fledged Generalised Linear Mixed Model for changes in abundance of prey and number of rough grassland dominated quadrats General Linear Model for weather variables and sheep abundance
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐++ ☐+	Comments: p-values given

Were confidence intervals and or p-values for the effect estimates given or calculable?	☐- ☐NR ☐NA	
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: A number of subjective/proxy measures but generally robust
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: Influence of grazing levels on vegetation type and structure (& therefore vole numbers) will depend on original vegetation, this is not reported Findings may not be applicable on sites with grazing by wild herbivores

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Anderson, P & Yalden, D. W.
	Year	1981
	Aim of study	To quantify the change in moorland vegetation and sheep statistics, and discuss the significance of the changes for red grouse and other wildlife
	Study design	3
	Quality score	+
	External validity	
Population and setting	Source population	Northern Peak District moorland
	Eligible population	Area previously mapped (1913) by Moss
	Inclusion and exclusion criteria	As above
	Setting	Moorland in 6 parishes of northern Peak District

Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	Agricultural grazing levels and change since 1930s
	Control/comparison description	N/A
	Sample sizes	Area wide survey
	Baseline comparisons	1913 Moss map, re-mapped at 1:25000 and areas of different heathland vegetation measured
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Change in extent of heathland vegetation communities
	Secondary outcome measures	Effect on red grouse numbers and other moorland bird species
	Follow-up periods	Change over a 66 year period for vegetation and 49 year period for sheep numbers
	Methods of analysis	Estimates of change. No statistical testing
Results		Mapping suggests a net loss f 56km^2 of heather to 1979, being 64% of its former extent. Heath has generally been replaced by grassland dominated by wavy hair-grass and mat-grass with bilberry. The 'grassland with much heather' on the summit of bleaklow has been replaced by crowberry with bilberry, mat grass, wavy hair-grass and heath rush. Of bilberry heath on rocky slopes and ridges, 46% had been lost. There had however been expansion of heather into cotton-grass areas aided by artificial drainage and gully

		erosion. Anecdotal evidence suggests most vegetation change took place since the 1930s. Sheep numbers overall increased by about four times between the 1930s and 1970s. The pattern holds at the parish level. The average stocking rate was 2.07 sheep ha^{-1} in 1977 compared to 0.7 sheep ha^{-1} in the 1930s. Bilberry appears to withstand moderate sheep grazing better than heather, but often reduced to a short dense form. Possible impacts from habitat change include loss of golden plover from some breeding sites and a reduction in mountain hare numbers. Estimates of loss of grouse are put between 85 000 and 118 000, depending on approach taken (grouse bags, habitat loss). A decrease in gamekeeping and shepherding is reported. Wheatear may have benefitted from increased grassland, but twite, ring ouzel and emperor moth may have suffered.
Notes	Limitations identified by author	Uncertainty of accuracy of original map, possible differences in interpretation of communities and crude level of reporting of sheep numbers
	Limitations identified by review team	No statistical analysis of correlative or causal relationships
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	

Quality Assessment Checklist: Qualitative Study v2.0

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ____Moorland Grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Anderson, P. & Yalden, D. W. (1981). Increases sheep numbers and the loss of heather moorland in the Peak District, England.
Study Design Category	3
Assessed by & when	D Martin 6/12/12

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?	☐ Appropriate	Comments: Semi-quantitative approach change in heather areas derived from mapping, and sheep numbers from June census data.
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: to quantify change in vegetation and relate to sheep numbers
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	☐ Defensible	Comments: Paper based mapping and comparisons. Pre-dates computerised mapping. The basis for the study was the existence of the original 1913 veg map by Moss

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?	☐ Defensible	Comments:

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: repeat mapping from aerial photographs and field surveys.

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?	☐Clearly described ☐Unclear ☐ Not described	Comments: Not relevant

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:

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 Rigorous	Comments: Mapped areas of each vegetation type compared and percent change measured. No correlation analysis with change in sheep numbers, due to limitations in the data.
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: Context well described.
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: Stock numbers at the parish level may not accurately reflect change on the moorland. Various external factors may affect reporting of sheep numbers. However the same trend is seen over a number of parishes outside of this study. The accuracy of the original map could be questioned, and subsequent interpretation of vegetation classes.
5.4 Are findings convincing? For example:	☐ Convincing	Comments:

-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?	☐ *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?	☐ Adequate	Comments:

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?	☐ Appropriately	Comments:

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?	☐ +	Comments:

- Was the data collection and record keeping systematic?		

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

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ___Moorland Grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? f) What factors influence spatial patterns of grazing? How effective are tools such as shepherding and burning in influencing grazing distribution, and how do they interact with stocking rates to achieve improvements in habitat condition and ecosystem services?
Study Citation	Anderson, P. & Radford, E. (1994) Changes in vegetation following reduction in grazing pressure on the National Trust Kinder Estate, Peak District, Derbyshire, England. Biological Conservation, 69, 55-63
Study Design Category	2
Assessed by & when	D Martin 12/12/12

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: Since it is a monitoring/ case study approach, source, eligible and sample area are the same. Vegetation, extent of erosion and recent change is broadly 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?	☐++	Comments: Eligible area as per source. Broad vegetation types given.
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: Monitoring was targeted at partly or completely bare ground, to monitor restoration – subjectively sampled. Twelve permanent transects established.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: Whole area subject to reduced grazing through shepherding. No estimate of grazing pressure in vicinity of sample areas. No comparison/ control.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Average grazing pressure calculated for whole area based on sheep gather numbers. No estimates of local grazing pressure in the sample area.
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: Could be confounded by climatic, in rainfall, slope and surface erosion and other grazing animals e.g. hares.
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: Vegetation cover and substrate measured on 21 pin hits at 1m intervals on a 10 or 12m transect. Biomass from small quadrat samples collected at random within nearby vegetation communities. Twenty-five samples cut down to 10 by discarding lightest and heaviest samples, due to resources. Reproductive capacity (flowering) of D flex and V m from small random quadrats. There was a nearby ungrazed control for comparison.
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, although biomass sample reduced in number, but probably adequate.
3.3 Were all important outcomes assessed? Were all important positive and negative	☐++	Comments: Yes, simple study to identify levels of colonisation.

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: Eight year study. Will identify trends but only follows early stages of colonisation and succession.

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:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: Only broad-scale changes in sheep numbers considered
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: t-tests to detect botanical differences over time, and ANOVA for flowering and Spearman's Rank Correlation in comparing with ungrazed areas. Biomass samples could not be analyses statistically.
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: Three levels of p given for t-test and anova.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐-	Comments: Subjective sample, likely to be subject to bias.

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: Case study, specific shepherding conditions and not easily presented in terms of grazing pressure for comparison or implementation elsewhere.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? f) What factors influence spatial patterns of grazing? How effective are tools such as shepherding and burning in influencing grazing distribution, and how do they interact with stocking rates to achieve improvements in habitat condition and ecosystem services?

Study details	Authors	Anderson, P. & Radford, E.
	Year	1994
	Aim of study	To examine vegetation change (and colonisation of exposed peat) following reduction in grazing pressure on Kinder plateau.
	Study design	2
	Quality score	-
	External validity	-
Population and setting	Source population	Since it is a monitoring/ case study approach, source, eligible and sample area are the same. Vegetation, extent of erosion and recent change on the plateau is broadly described.
	Eligible population	Eligible area as per source. Broad vegetation types given.

	Inclusion and exclusion criteria	Monitoring was targeted at partly or completely bare ground, to monitor restoration – subjectively sampled. Twelve permanent transects established.
	Setting	Kinder plateau, Peak District, Derbyshire. Transects located on sloping north to west facing slope between 450m and 530m.
Methods of allocation to intervention/control	Methods of allocation	Case study – sheep reductions at scale of moorland unit. Study transects targeted at eroded areas.
	Intervention description	Whole area subject to reduced grazing through shepherding. No estimate of grazing pressure in vicinity of sample areas, only average pressure over whole area based on counts at gather. No comparison/ control.
	Control/comparison description	No control, although flowering was compared with an un-grazed road cut.
	Sample sizes	12(ultimately 10 relocated) transects of 10-12 pin frames with 21 pins.
	Baseline comparisons	Vegetation and biomass measurements in first year of sheep reductions.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Change in vegetation cover and biomass
	Secondary outcome measures	Flowering (reproductive capacity)
	Follow-up periods	Eight-year study
	Methods of analysis	t-tests to detect botanical differences over time, and ANOVA for flowering and Spearman's Rank Correlation in comparing with ungrazed areas. Biomass samples could

		not be analyses statistically.
Results		At the start of the study bare ground had a mean percentage frequency of 51%, with *D flexuosa* the most abundant species at 41%. The steepest unstable slope had the most bare ground (64%) and lowest *D flexuosa* cover (30%). Other species were at very low cover with suppressed heather and cotton grass species all at less than 1% cover on average. It was estimated that sheep grazing pressure reduced from 2.5 ewes ha^{-1} in 1882 to between 0.18 and 0.43 ewes ha^{-1} over the course of the study. Over the study *D flexuosa* increased linearly to 83%, from both vegetative spread and seed, mirrored by a reduction in bare ground. After an initial lag effect *Calluna* was seen to spread from both established plants and seed, particularly on gently sloping mineral soils where it attained a cover of 32% from less than 1% initially. *Calluna* did not colonise the steepest slope, but *D flexuosa, N Stricta* and *V myrtillus* all spread. Above ground biomass of D flexuosa changed little initially but increased markedly after 4 years, to six times the initial mean biomass. New growth of Vaccinium increased four-fold between 1983 and 1988, and there was a rapid increase in biomass of old growth between 1985 and 1986. A decline in both fractions was observed in 1990, which may be down to climatic effects. The results show that stock reductions to low levels (below 0.5 ewes ha^{-1}) allows vegetation to recolonise mineral and peaty soil. It is however slow, at least at altitude (c 500m), taking eight years to increase from mean cover of 49% to 92%. This cover was however achieved in only five years on lower slopes on mineral soil, but much bare ground remained on steep slopes. *D flexuosa* spread rapidly from vegetative growth and seed, but heather and bilberry continued to spread and the community may change to *Calluna-Vaccinium* in the longer term.
Notes	Limitations identified by author	

	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Baines, D
	Year	1996
	Aim of study	To investigate relative importance of predator numbers through the role of gamekeeper and habitat quality related to grazing intensity, on density and breeding success of black grouse
	Study design	2
	Quality score	-
	External validity	++
Population and setting	Source population	Source population is the extent of UK moorland habitat within the range of black grouse. Only covered in very general terms
	Eligible population	The eligible population is moorland occupied by black grouse, grazed by sheep and/ or red deer, and/ or managed for red grouse through burning and predator control. Again

		described broadly, and likely to be representative of source area, but some expansion on characteristics of lightly grazed sites
	Inclusion and exclusion criteria	Sites selected to fulfil treatment description. Characteristics of area, sheep numbers, keeper numbers are given for each site.
	Setting	Five moorland blocks in N Pennines, The Scottish Borders, Central Perthshire, North Perthshire and Speyside
Methods of allocation to intervention/control	Methods of allocation	Exposure was pre-existing grazing and keepering conditions, but selected to represent each of four combinations within a geographical block. Paired moors (low and high grazing?) within blocks were located adjacent to each other where possible to minimize differences in soil, geology etc.
	Intervention description	Intervention is light or heavy grazing with sheep and/or deer (not well specified) in combination with predator control (gamekeeper present) or none (no gamekeeper)
	Control/comparison description	Heavy grazing with no gamekeeper could be viewed as the control
	Sample sizes	Each of four treatment combinations applied in five blocks
	Baseline comparisons	No baseline as such as a comparative study
	Study sufficiently powered	Not reported
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	key outcomes are black grouse male and female densities and brood numbers
	Secondary outcome measures	Breeding success (number of young reared per female)

significance)	Follow-up periods	All sites surveyed for birds in each year 1991-93. However as treatment and control is the pre-existing management the conditions will have been in place for different durations.
	Methods of analysis	Largely analysis of variance techniques (anova and manova) and comparison of means. Block (geography) effects were tested and differences between years in breeding success.
Results		Moors with higher grazing intensities had on average 32% shorter and 36% less vertical vegetation cover. No significant effect on species composition (DM note: may be due to coarse measures of composition used). Heavily grazed moors supported 41% fewer invertebrates, with some key groups (Lepidoptera, Araneae, Hemiptera) less well represented. Black grouse breeding success differed between years and regions, but also between management treatments, being 37% lower on heavily grazed moors. The lower level of grazing allows the development of ground cover which is correlated with higher numbers of preferred invertebrate food, which reduces the need for large movements in foraging broods. The higher success on lightly grazed moors was independent of the presence of a gamekeeper. Gamekeeper presence was not associated with higher breeding success despite there being three times fewer carrion crows on keepered moors. Tall vegetation may aid survival in situations where numbers may otherwise be severely reduced by predators. Estimates of 1.5 -2 chicks per year necessary to maintain a stable population were attained on the lightly grazed moors, but not on the heavily grazed moors.
Notes	Limitations identified by author	Lack of data on mammalian predators, limited time spent on estimating carrion crow numbers. In this extensive (rather than intensive) study it is not possible to determine the main cause or stage of breeding failure and identify relationships with habitat quality.

	Limitations identified by review team	Weak association of measured variables with actual sheep and deer density, and limited data on keepering effects (e.g. verying effort could be a confounding factor).
	Evidence gaps and/pr recommendations for further research	Futher study of the relationship between increased density of large herbivores and abundance of black grouse. To be examined experimentally by manipulating
	Sources of funding	Dulverton Trust, EN, Scottish Forestry trust, SNH, WWF

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

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Baines, D. (1996). The implications of grazing and predator management on the habitats and breeding success of black grouse. Journal of Applied Ecology, 33, 54-62
Study Design Category	
Assessed by & when	D Martin 18/10/12

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: Source population is the extent of UK moorland habitat within the range of black grouse. Only covered in very general terms
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: The eligible population is moorland occupied by black grouse, grazed by sheep and/ or red deer, and/ or managed for red grouse through burning and predator control. Again described broadly, and likely to be representative of source area, but some expansion on characteristics of lightly grazed sites
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: Inclusion/ exclusion criteria are clear, selected to fulfil treatment description. Characteristics of area, sheep numbers, keeper numbers given for each site. Sites chosen to meet treatment criteria. Not clear how they were initially identified and the extent of initial choice – may be sources of bias.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: Exposure was pre-existing grazing and keepering conditions, but selected to represent each of four combinations within a geographical block. Paired moors (low and high grazing?) within blocks were located adjacent to each other where possible to minimize differences in soil, geology etc.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐-	Comments: Explanatory variables are prey (invertebrate) availability, grazing pressure, with surrogate measures of sward height and density (visibility of chequer board through sward). The number of vegetation measurements may not adequately represent grazing impacts on the site. No estimates of grazing animal density or impact, and grazing levels not well specified. Crows were counted as a predator, but no estimates of mammalian predators.
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: Treatments are existing levels of grazing and keeper activity, so no scope for contamination as such, but in reality sites will a spread of grazing pressures and keeper effects.
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: Paired sites adjacent where possible to minimise climate, soil and geology effects. Brood counts made away from the boundary to minimise edge effects.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes – range of sites through northern uplands within the range of black grouse and typical of management of upland moorland areas

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: Main outcome measures were counts of displaying males from maximum count at two lek visits. Breeding success estimated from dog searches for broods in a representative area, adjusted to locate 10 females. Densities of males and females are presented. Indices of breeding success (young reared per female, percentage of females with broods) were calculated if five or more females located, and mean brood size based on three or more broods. Habitat is assesses via vegetation composition – this is

		done through a fairly crude measure of cover of 6 broad categories of plant species
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: It seems that all black grouse counts were completed. Although maybe more related to explanatory variables, crow numbers were not counted consistently in each year, and invert/ veg measurements appear to have been made only in first year.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Yes – key outcomes are black grouse numbers and breeding success.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: Yes – particularly bird measures. Some questions over the relevance of the vegetation cover estimates.
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	Comments: All sites surveyed for birds in each year 1991-93. However as treatment and control is the pre-existing management the conditions will have been in place for different durations.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Again it is not known how long the grazing and keepering levels have been in place, but probably safe to assume general management has been in place long enough for effects to develop.

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 given. Five 'replicates' of each treatment combination, but from different geographic areas.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: Limited number of explanatory variables used, basically just presence/ absence of keepering and two grazing states. May have benefitted from more detailed exploration of grazing levels and habitat characteristics
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐+	Comments: Largely analysis of variance techniques and comparison of means. Block (geography) effects were tested and differences between years in breeding success.

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?	☐++	Comments: p values of anova and manova given
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: Not clear how sources of bias are minimised, as basis of site selection not fully explained. Likely to be confounding factors not fully adjusted for, but difficult in this type of study. Sites chosen to represent geographical range of black grouse, and paired to minimise environmental effects. However some concerns that the links between vegetation measures and grazing is weak, as are the measure of predator impacts and possible variable effort in gamekeeping.
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: Blocks located throughout range of black grouse, and broad habitats and management is typical of the general management of wider area.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	BARDGETT, R. D., JONES, A. C., JONES, D. L., KEMMITT, S. J., COOK, R. & HOBBS, P. J
	Year	2001
	Aim of study	To study successional transitions to determine how variation in the history and intensity of grazing alter the biomass, activity and structure of the soil microbial community
	Study design	2
	Quality score	+
	External validity	++
Population and setting	Source population	The source population is said to be the UK upland ecosystem. Below the tree line.
	Eligible population	The eligible population is successional areas from three biogeographic zones, selected to cover a range of history and intensity of management.
	Inclusion and exclusion criteria	Sample areas chosen to represent a range of grazing pressure. Chosen to be typical in terms of geology and drift, soil base status etc. The transition covers oak-birch woodland, heather moorland, *Nardus* and *Agrostis-Festuca* habitats.

	Setting	Range of habitats representative of different grazing pressures in Snowdonia, Lake District and Yorkshire Dales. Mainly 250-300m in altitude.
Methods of allocation to intervention/control	Methods of allocation	Site selection likely to have been subjective to meet successional criterion, i.e. to achieve the range of grazing histories. Within each treatment, three randomly located replicate plots were identified, to reduce errors from pseudo-replication
	Intervention description	A table describes the vegetation type and prevailing management at each site. The transitions cover six treatments from permanently ungrazed (oak-birch woodland and heathland) to heavily grazed (8-16 ewes ha^{-1} yr^{-1} on *Agrostis – Festuca* grassland).
	Control/comparison description	Long term ungrazed (woodland) could be considered control/ comparison
	Sample sizes	Three locations with each of 6 treatments. Three replicate plots within each treatment, and Ten soil cores taken in each replicate.
	Baseline comparisons	Sites necessarily vary in their starting point. Chosen to be as similar as possible in soils etc. No detailed veg data at start, only broad descriptions
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Soil C, C:N ratio, Microbial activity – PLFA composition
	Secondary outcome measures	Fungal: bacterial ratio, nematode analysis
	Follow-up periods	All sampled in June 1989. Treatment exposure necessarily varies in time between treatments – different lengths of grazing exclusion.

	Methods of analysis	ANOVA within biogeographic locations to determine variance attributed to the grazing gradient. Fishers PLSD to test fro between mean differences. Also ANOVA with site as a replicate to account for pseudoreplication.
Results		Significant trends in soil C were seen at the Lake District and Snowdonia, being highest in the lightly grazed and short-tem ungrazed sites, and lowest in heavily grazed grassland. Soil C:N ratios was affected by grazing influence at all locations, being highest in the long-term ungrazed grasslands. Soil pH showed a general trend of increasing acidity with reduced grazing pressure from the heavily grazed to lightly grazed grassland. Microbial biomass, as measured by total phospholipid fatty acid (PLFA), varied significantly along the gradient, and was highest in the lightly grazed treatment at each site, declining along the gradient to long-term grazing exclusion. The ratio of fungal to bacterial PLFA ratio varied significantly along the gradient for all locations, and generally highest in moderately grazed grassland. PLFA evenness, a measure of the relative distribution of microbial PLFAs, tended to decrease from the ungrazed and lightly grazed treatments to the heavily grazed. The data shows there are consistent broad scale trends in soil microbial communities along successional gradients that are related to grazing intensity. Microbial biomass is greatest at low to intermediate levels of grazing and evenness (i.e. lack of dominance of individual groups) declines as grazing intensity increases. This evidence suggests that decomposer-related processes, such as nutrient cycling, may be optimal at intermediate grazing levels. This was not fully supported by soil respiration rates however, which was highest in lightly grazed treatments at only one site. There was evidence that intensively grazed sites were dominated by bacterial based decomposition, whilst in the lightly grazed or ungrazed treatments fungi have a proportionately greater role.
Notes	Limitations identified by author	Lack of effect of grazing intensity on total nematodes may mask the response of different nematode groups. Lack of data on plant productivity of soil process to relate to the findings on soil

		microbial patterns.
	Limitations identified by review team	Grazing levels largely inferred from vegetation type. NO estimates of grazing pressure given at the site level, or used in analysis. The link between grazing levels and dominant veg type is assumed.
	Evidence gaps and/pr recommendations for further research	Further work suggested to establish whether identified trends are temporally robust and to determine significance of these changes in relation to soil-level ecosystem processes of decomposition and nutrient cycling.
	Sources of funding	

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	BARDGETT, R. D., JONES, A. C., JONES, D. L., KEMMITT, S. J., COOK, R. & HOBBS, P. J. 2001. Soil microbial community patterns related to the history and intensity of grazing in sub-montane ecosystems. *Soil Biology & Biochemistry,* 33**,** 1653-1664.
Study Design Category	2
Assessed by & when	D Martin 18/01/12

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: The source population is said to be the UK upland ecosystem. Below the tree line.
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: The eligible population is successional areas from three biogeographic zones, selected to cover a range of history and intensity of management.
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: Sample areas chosen to represent a range of grazing pressure. Chosen to be typical in terms of geology and drift, soil base status etc. The transition covers oak-birch woodland, heather moorland, *Nardus* and *Agrostis-Festuca* habitats.

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: Site selection likely to have been subjective to meet successional criterion, i.e. to achieve the range of grazing histories. Within each treatment, three randomly located replicate plots were identified, to reduce errors from pseudo-replication
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: A table describes the vegetation type and prevailing management at each site. The transitions cover six treatments from permanently ungrazed to heavily grazed.
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, particularly the ungrazed treatments. Not always clear how long grazing treatments have been as 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?	☐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?	☐+	Comments: Not reported, although as some sites part of farming systems there may be other interventions that have affected some treatments at some sites.
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: Some of the treatments reflect typical grazing management for the habitat types.

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: Standard soil analysis techniques, soil microbial communities by PLFA – accepted technique
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: All sampled in June 1989. Treatment exposure necessarily varies in time between treatments – different lengths of grazing exclusion.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐NA	Comments:

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: Sites necessarily vary. Chosen to be as similar as possible in soils etc.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?		Comments:

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	
4.3 Were the estimates of effect size given or calculable?	☐NR	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: ANOVA within biogeographic locations to determine variance attributed to the grazing gradient. Fishers PLSD to test fro between mean differences. Also ANOVA with site as a replicate to account for pseudoreplication.
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:
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: Vegetation type taken for a surrogate for grazing pressure and history. Grazing not controlled
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:

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

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	
Study Citation	Britton, A.J., Pearce, I.S.K. & Jones, B. (2005) Impacts of grazing on montane heath vegetation in Wales and implications for the restoration of montane areas. Biological Conservation 125, pp512-524
Study Design Category	
Assessed by & when	D Martin 10/10/12

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: Limited description of the habitat and UK extent, some history of change in habitat and environmental conditions in Wales.
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: Location and NVC types given
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: Sample area is most extensive area of the habitat in Wales. Samples located at random, stratified by three communities. Min distance of 20m from footpath. May be a degree of subjectivity in selecting sample communities, and defining comparisons (degraded). Not entirely clear if stratification took place before sample location. Or community types attributed after. Samples described in NVC terms and key dominants.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Relative grazing pressure measured via dung counts. May be some bias as the communities have some altitudinal separation, and altitude may influence grazing pressure/ pattern
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes, based on observations of change in sheep numbers and atmospheric deposition over preceding 40 year period. Atmospheric deposition not measured directly, but combined effects of nutrient impacts measured via soil and tissue chemistry.
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: Not treatment based, but measuring prevailing environmental conditions
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: As the communities are loosely distributed by altitude (Vaccinium community generally lower than Racomitium) there may be other factors that vary with altitude not measured in the study. Whilst the sampling was designed to avoid major variations due to altitude, there may be large variations in for example soil chemistry over a short range. Sample done in two groups – two months apart
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes, but very restricted habitat, which may vary throughout its geographic range

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: Dung counts are a surrogate measure of grazing pressure, may be factors which influence dunging rate in different places. Soil and tissue chemistry likely to be reliable. Cover by estimation – poss observer inconsistency.
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? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: Dung counts are a surrogate measure of grazing pressure, may be factors which influence dunging rate in different places
3.5 Were there similar follow up times in exposure and comparison groups?	NR ☐NA	Comments: Survey rather than treatment approach
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Survey measures the results of long-term exposure to grazing an N deposition.

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. Suspect power may be quite low for some variables
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments: Yes, grazing pressure plus range of soil chemistry variables
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: Ordination techniques used to compare composition and environmental variables. Anova used to identify sig differences in soil and tissue chemistry between veg groups
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: P values given for regression of dung counts and tissue content, and differences in soil and plant chemistry variables

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: Survey type rather than treatment approach. Random sampling to reduce bias, but relatively small sample size.
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: This habitat is limited nationally, and no similar studies elsewhere. However

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland Grazing
Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Britton, A.J., Pearce, I.S.K. & Jones, B
	Year	2005
	Aim of study	To investigate the links between species composition and grazing impacts, current condition of montane heath in Wales and whether a reduction in grazing likely to be sufficient for restoration.
	Study design	
	Quality score	-
	External validity	+
Population and setting	Source population	Montane heath habitats
	Eligible population	Montane heath dominated by Vaccinium and Racomitrium in a mountain ridge in Snowdonia, Wales

	Inclusion and exclusion criteria	Above 870m, min distance 20m from trampled path
	Setting	Close to top of extensive mountain ridge in Snowdonia
Methods of allocation to intervention/control	Methods of allocation	Randomly allocated, stratified by broad vegetation type. May be subjectivity in
	Intervention description	Intervention is the prevailing grazing conditions, assessed by surrogate measure of dung counts
	Control/comparison description	Not a treatment approach. Comparative study - relative grazing levels on the three vegetation types compared.
	Sample sizes	Total 37 6mx 6m plots. Group sizes 11-14
	Baseline comparisons	One- off survey
	Study sufficiently powered	No/ not reported –likely that power is low to detect significant change in some variables
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation composition and cover, Soil chemistry toal C, N, P, pH (H+), exchangeable cations (Al, Ca, Fe, K, Mg, Mn, Na)
	Secondary outcome measures	Plant tissue chemistry of key species (total N, P)
	Follow-up periods	One – off survey
	Methods of analysis	Ordination techniques used to compare composition and environmental variables. Anova used to identify sig differences in soil and tissue chemistry between veg groups

Results		Sheep occupy Vaccinium areas early in summer, but habitat use evens out later in summer. Soil profile under degraded vegetation is indicative of lost organic horizons (lower C, N, P, slightly higher pH). The three species studied (F ovina, C, bigelowii and V myrtillus) differed significantly in nutrient concentrations and N:P ratios, the two graminoids being P limited, and Vaccinium N limited. Comparison with other studies suggest the site is exposed to high N deposition from the atmosphere, or dung and urine. Tissue N of Vm was positively associated with dung deposition. Ordination suggests an association of current vegetation with soil properties and altitude. In particular high soil pH associated with degraded soil, and altitude the main association with Vm and Rl. No strong association between dung deposition and habitat category. It is likely that habitat degradation occurred rapidly in early years of increased grazing (see Welch, 2005 – Racomitrium response to increased grazing pressure). Degradation likely to be on-going however, as dung counts suggest high grazing pressure. Overall results show loss of organic horizon in most severely degraded vegetation, so physical and chemical conditions likely to be unfavourable for re-colonisation – sp restoration likely to be slow. Restoration perhaps best targeted where characteristic species still present.
Notes	Limitations identified by author	
	Limitations identified by review team	Relatively limited study, geographically and in terms of sample size and location at the site
	Evidence gaps and/pr recommendations for further research	Experimental work on the effects of reduction or removal of grazing and the ability of species to re-colonise vegetation in different states of degradation.
	Sources of funding	CCW, SERAD, NERC

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Calladine, J., Baines, D. & Warren, P.
	Year	2002
	Aim of study	To investigate the effects of reduced grazing (through agri-environment schemes) on population density and breeding success of black grouse
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	The source population is the extent of moorland and moorland fringe habitat in the North Pennine range of black grouse
	Eligible population	The eligible area is where black grouse are known to occur are likely to be broadly r
	Inclusion and exclusion criteria	Black grouse present, recent reductions in grazing at the treatment sites
	Setting	North Pennines, Northern England

Methods of allocation to intervention/control	Methods of allocation	Opportunistic – where AE schemes implemented at sites known to the Black Grouse recovery Project.
	Intervention description	Year round reduction in sheep grazing on all or part of the study area (on average 1.1 sheep ha^{-1} summer compared with 2.4 at reference site, and 0.5 winter compared with 1.7)
	Control/comparison description	Comparison with paired sites with no reduction, but typical farm stocking rates. Minimum of 5 km between paired sites, but reasonably close (mean 9.3km)
	Sample sizes	10 treatment and 10 reference.
	Baseline comparisons	No baseline – comparative study – treatments in place pre-study.
	Study sufficiently powered	N o power calculation presented. Sample size judges to be adequate for this type of comparison study.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Number of displaying males at lek, female population density and brood density.
	Secondary outcome measures	Breeding success
	Follow-up periods	Follow up time varies between treatment sites. Length of time since stock reduced vary between sites (period of treatment prior to baseline counts varies from 1 to 5 years).
	Methods of analysis	
Results		No significant difference in proportion of occurrence of vegetation communities between treatment and reference sites, but generally taller mean sward height in treatment areas and reduced variation in sward height.

		Displaying males at leks showed a significantly different trend with an average increase of 4.6% at treatment sites, and reduction of 1.7% at reference sites. There was a similar but non-significant difference in trends in female birds observed – the effect reduced in summer as females appear to avoid tall swards for breeding. There appears to be biggest positive difference in trend co-efficient of females where treatment area (sward <30cm) is 100ha or less. A marginally non-significant relationship of trend in number of displaying males with time since grazing reduced is reported, with apparent peak at 5-7 years. This is however weak. A higher percentage of females (54%) at treatment sites had broods than at reference sites (32%). There was no difference in brood size. Brood size was consistently greater at treatment sites, but not significant in every year. There was no apparent relationship between breeding success and area, grain or age of grazing restriction.
Notes	Limitations identified by author	No attempt made to determine 'optimal' stocking densities. Some key food species would be under-sampled by the methods, but may have beneficial influence. Small brood sample size (mean of 1.6 per site)
	Limitations identified by review team	Limited range of vegetation measures, and lack of quantification of sheep stocking densities per site (only means for each treatment given).
	Evidence gaps and/pr recommendations for further research	Need to further elucidate the mechanisms of the effect of reduced grazing on black grouse numbers, and whether the effect can become limiting. Further investigation of temporal change in the influence of grazing reduction is required, and whether conditions start to deteriorate after a period of time. Further understanding of landscape-scale dynamics, and impacts of wide-scale implementation of similar schemes.
	Sources of funding	Part of monitoring programme of North Pennine Black Grouse Recovery Project, funded by English Nature, The Game Conservancy Trust, MoD and RSPB. Supplementary support from National Wind Power

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

Name of Evidence Review: _____Uplands_____

Name of Review Sub-topic (if any): _____Moorland Grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Calladine, J., Baines, D. & Warren, P. (2002) Effects of reduced grazing on population density and breeding success of black grouse in northern England. Journal of Applied Ecology 39, 772 -780
Study Design Category	2
Assessed by & when	David Martin 17/10/12

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 source population is the extent of moorland and moorland fringe habitat in the North Pennine range of black grouse. Briefly described in terms of altitudinal range and broad vegetation types.
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: The eligible area, i.e. where black grouse are known to occur are likely to be broadly representative of the wider habitat, but vary in key attributes which increase the suitability for the 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?	☐-	Comments: The basis of selection of the twenty study areas is not clear – likely to be subjective or opportunistic, based on sites the Black grouse project have involvement. Also will not include full range of vegetation condition as sites have been selected for agri-environment restoration. Two criteria were however applied: occupation by black grouse and recent sheep reductions on at least part of area. Min distance of 5km between paired samples was applied.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: The exposure is a reduction in grazing (magnitude unspecified) through agri-environment schemes. The area of reduced grazing, proportion of the study site affected, and length of time since stock reduced vary between sites (period of treatment prior to baseline counts varies from 1 to 5 years). Number of paired sites reasonable for comparative study.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Explanatory variables based on theoretical need for vegetation structure for breeding black grouse. However, only very simple measure of vegetation height used, and counted in one year, and limited attempt to quantify grazing levels.
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: No contamination reported (i.e. comparison sited subject to similar grazing reductions). However, there is no control over grazing levels at comparison sites, and they will vary between sites and possibly over time.
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: Confounding factors include sporting management (8 treatment, 7 reference) where predators are controlled. Avian predators were estimated, but not considered further and affects assumed to be similar between the two groups, but this is not known. Also may be other structural elements, possibly related to length of period of grazing reduction, that have effect.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes – moorland and upland fringe habitats likely to be fairly typical of upland areas in the black grouse range.

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: Outcomes are observed presence of displaying males at leks. Females and broods counted by systematic searches using dogs. May be a degree of subjectivity, but experienced surveyors used.
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: Not all sites surveyed for birds in first year (1996). All sites done in subsequent years 1997 – 2000.

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Yes – various measures of bird density and breeding success made.
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: There is some variation in the number of sites assessed in first year of bird counts – leks counted at four treatment and seven reference sites, and hens at eight treatment and five reference sites in 1995. Treatments (reduced grazing) have been in place for variable amounts of time.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Follow up time varies between treatment sites. Likely to be long enough to detect some effects of reduced grazing, but not long-term effects? Sites will be at different stages of transition, however this is taken into account in analysis.

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 power calculation presented. Sample size judges to be adequate for this type of comparison study.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: Additional explanatory variables of time since reduction and proportion of area covered were considered in analysis, but vegetation variables very limited – e.g. no measures of spatial heterogeneity.
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: Analysis took account of differences in time and area of reductions, by comparing the differenced in population trend coefficients between treatment and reference pairs, and differences in female/ brood ratios. Additionally, 'grain' assessed – edge/ area ratio.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐++	Comments: Significance levels (p value) and standard errors given for all analyses.

Were confidence 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. 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: May be unintentional sources of bias of confounding effects in site selection – as subjective/ opportunistic. Will encompass degree of variability in environmental factors, and other management including sporting/ predator control – assumptions made. Attempts made to adjust for variation in area and duration of grazing in the analysis.
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: Sites likely to be fairly representative of existing black grouse range (present at all sites at start of study), and encompass the range of environmental conditions. However since sites are agri-environment restoration, will have been selected as sub-optimal habitat condition, particularly low dwarf shrub cover.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): ____Moorland grazing_____

Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? What factors influence spatial patterns of grazing? How effective are tools such as shepherding and burning in influencing grazing distribution, and how do they interact with stocking rates to achieve improvements in habitat condition and ecosystem services?
Study Citation	Clarke, J.L., Welch, D. & Gordon, I.J. (1995) The influence of vegetation pattern on the grazing of heather moorland by red deer and sheep I. The location of animals on grass/ heather mosaics. Journal of Applied Ecology 32. 166-176 *and:* II The impact on heather. Journal of Applied Ecology 32, 177-186
Study Design Category	2
Assessed by & when	D Martin 26/10/12

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: Population is the UK extent of heather moorland and heather/ grass mosaics. Not described in 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: The experimental sites had varying-sized patches of grass occurring in a matrix of heather. Representative of dry heath communities, but wet heath and bog not considered.
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: Experimental sites chosen to represent different patterns of grass distribution, in a natural mosaic. Site selection therefore subjective. One site manipulated by introducing sown grass patches, due to concerns over possible confounding of naturally occurring grass patches. Background vegetation typical of species poor heather dominated vegetation, and well within altitudinal range.

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: Only two replicates per treatment, not applied randomly. Each treatment consisted of creating grass patches such that the grass: heather ratio was always 1:5, but in 1 large, 4 intermediate or 12 small patches. Unsure why intermediate grass plots are l-shaped rather than rectangular – reduces the amount of heather between grass areas. Each plot subject to alternating periods of grazing with sheep and deer at equivalent LUs. Adjacent plots were grazed by different species at any one time, as same species tend to rest along a shared fenceline. A second experiment compared two stocking rates of sheep on the three patch size treatments, again alternating over three grazing periods. Also a preliminary experiment between July and November 1991. Three group sizes of sheep and three of deer rotated around each plot for 6 grazing periods, with 3 week break between periods 3 and 4.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Grazing treatments well described and tabulated, with diagram of layout, so could be replicated. Animal groups kept together and rotated around plots. On an annual basis stocking rates are typical of farm practice, but concentrated in time and space for the purposes of the study. On heather utilisation study the higher sheep stocking rate (22 ewes) probably at high end of typical annual grazing pressure.
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: Exposure periods relatively short (three bursts of 14 days), but likely to be adequate to allow spatial grazing patterns to be observed. No problems reported with implementation. Grazing treatments only carried out in one season, so may not take account of weather factors that might affect grass growth. Heather utilization study longer (July- Nov).
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: There is no control/ comparison as such, but comparing three vegetation states. No apparent contamination.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups?	☐++	Comments: None apparent.

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.	☐+	Comments: Likely to be representative of dry heath/ grass mosaics, but not other moorland communities. However, experiment involved artificial manipulation of the vegetation.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The overall stocking rates calculated on an annual basis are broadly in line with practice, although possibly slightly high on highest grazing rate on heather utilisation study. However the small plots approach is not necessarily representative of normal ranging practice and spatio-temporal grazing patterns, but this has to be sacrificed to some extent in well controlled experiments.

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: observation and recording of animal location (grass patch, heather zones < 5m from grass; 5-30m and >30m) and activity – reasonably objective as should be clear which vegetation type animals are on. Vegetation heights measured objectively with standard HFRO sward stick. Heather utilisation in the 1991 study measured at 10 fixed points on a transect in permanent 5x20m quadrats. Ten quadrats per heather zone, except 4 plot/ zones where not enough space. Methods chosen to allow as many shoots as possible to be examined on two-day period. In the 1992 study utilisation measured at fixed points along transects in each zone. Proportion of shoots grazed has been shown to correlate well with more accurate measures of utilisation in terms or proportion of biomass removed (Armstrong & MacDonald 1992). Ten heights per quadrat also taken. Utilisation for a grazing period taken as the difference of proportion of shoots grazed at start and end of a grazing period, zone means multiplied by proportion of zone occupied by heather, and summed for plot mean.
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?	☐+	Comments: The heights of dwarf shrub surrounding grass patches (at different distances) were not

Were all important positive and negative effects assessed by the variables/measurements used?	☐-	measured. Location and behaviour information for utilisation experiment were lost in a fire.
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 – actual location of grazing animals, measured sward heights. Yes – measures of utilisation of heather is important in assessing likely grazing impact.
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: Only assessed over relatively short intervals for one season. Probably adequate to determine usage patterns, but longer-term effects not observed.

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: No comparison group as such, but comparing three vegetation states/patterns. Basic habitat type and species composition is similar. Utilisation levels and heather height measured at the start of each experiment, to allow for any spatial differences in previous utilisation. However in 1992 expt 2 the measured utilisation at end of period 1 was less than the expected starting value assumed to be the same as end of expt 1.
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?	☐NR ☐NR	Comments: No power analysis given, and lack of replication. Type of expected effects, but not magnitude, are set out. See comments below. Some analyses of effect quote low degrees of freedom, due to loss of orthogonality.
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments: Differences in proportion of animal sightings on different vegetation types during each grazing period, and mean percentages during different daytime periods given. Mean densities of grazing animals in the heather zone for each patch-size treatment and densities in the grass patches presented.
4.4 Were the analytical methods appropriate? Were any important differences in post-	☐++	Comments: Means and SE of measures calculated using Restricted Maximum Likelihood (REML) to account for non- orthogonality of experiment. Sources of variance calculated using Generalised

treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?		Linear Model (GLM). Regression analyses of factors affecting grazing time in the small patches. Effects of treatment, species and period analysed at plot level, using adjusted mean utilisation values to account for different sampling effort in zones. Effects of zone and interactions with zone used unadjusted data for grazed shoots. Similar techniques used as above. Because species and treatment effects estimated in both the plot and period strata there were insufficient degrees of freedom to fully estimate their effects. Results are quoted, but with lower confidence.
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: Standard errors given for mean values, p-values for regression equation R^2 values. Significance of F-values given for sources of variance.
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: Treatments were implemented well, but weak replication. The analysis took account of some of the limitations in experimental design. Only carried out over one season. Lack of power of some analyses due to experimental design. This is however a good example of a controlled grazing experiment. Changed to – on basis if reviewer QA exercise – limitations of not measuring night time grazing etc
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: Generalisable in terms of habitat – reasonably representative although obviously cannot represent the geographic and environmental variation in the habitat. The behaviour of livestock in small plots may be different from the open hill, and configuration of grass patches are artificial. However grazing choices are likely to be translatable and proportion of time spent and groupings of grazing animals likely to be translatable to other situations.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? What factors influence spatial patterns of grazing? How effective are tools such as shepherding and burning in influencing grazing distribution, and how do they interact with stocking rates to achieve improvements in habitat condition and ecosystem services?

Study details	Authors	Clarke, J.L., Welch, D. & Gordon, I.J.
	Year	1995 a&b
	Aim of study	To test the effects of size and distribution of grass patches on heather use by grazing animals, and the effect on animal distribution. To test the impact of sheep and deer on heather in terms of proportion of shoots grazed, and death of heather shoots, with respect to distance from grass patches.
	Study design	2
	Quality score	+ + - after QA exercise
	External validity	+ +
Population and setting	Source population	Population is the UK extent of heather moorland and heather/ grass mosaics. Not described in detail.
	Eligible population	The experimental sites had varying-sized patches of grass occurring in a matrix of heather. Representative of dry heath communities, but wet heath and bog not considered

	Inclusion and exclusion criteria	
	Setting	Mature heather moorland on east-facing slope of a hill at Glensaugh Research Station, NE Scotland.
Methods of allocation to intervention/control	Methods of allocation	Subjective – only two replicates of each patch size treatment – not randomised.
	Intervention description	Each treatment consisted of creating grass patches such that the grass: heather ratio was always 1:5, but in 1 large, 4 intermediate or 12 small patches. Unsure why intermediate grass plots are l-shaped rather than rectangular – reduces the amount of heather between grass areas. Each plot subject to alternating periods of grazing with sheep and deer at equivalent LUs. Adjacent plots were grazed by different species at any one time, as same species tend to rest along a shared fenceline. A second experiment compared two stocking rates of sheep on the three patch size treatments, again alternating over three grazing periods. Also a preliminary experiment between July and November 1991. Three group sizes of sheep and three of deer rotated around each plot for 6 week grazing periods, with 3 week break between periods 3 and 4.
	Control/comparison description	Comparing three configurations of grass patch, but no control site as such.
	Sample sizes	Two plots of each treatment, with half-hourly observations of animal activity and location. Vegetation measurements – species and height at 100 points in large grass patch, 50 in medium and 25 in each small patch. Heather utilisation in the 1991 study measured at 10 fixed points on a transect in permanent 5x20m quadrats. Ten quadrats per heather zone, except 4 plot/ zones where not enough space. Methods chosen to allow as many shoots as possible to be examined on two-day period. In the 1992 study utilisation measured at fixed points along transects in each zone.
	Baseline comparisons	Grass sward measurements made at start of each experiment. Measurements of

		utilisation and heather height made prior to start of each experiment.
	Study sufficiently powered	No power analysis given, and low replication. Paper states that changes in vegetation over the experiment means the three grazing periods could not be treated as simple replicates. Type of expected effects, but not magnitude, are set out. Because species and treatment effects estimated in both the plot and period strata there were insufficient degrees of freedom to fully estimate their effects. Results are quoted, but with lower confidence. Low confidence in baseline utilisation measures in expt 2 and discarding of period 1 measures reduced degrees of freedom available to detect differences in stocking rate effect.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Primary outcome measures are the observations of animal activity and occupancy of each patch and heather zones. Also measures of heather utilisation in each plot and zone in terms of proportion of shoots grazed.
	Secondary outcome measures	Animal group size (grouped animals considered to be <30m apart) and vegetation heights. Vegetation heights.
	Follow-up periods	Only assessed over relatively short intervals (3x 14 day periods for each experiment) for one season. Probably adequate to determine usage patterns, but longer-term effects not observed. Six short grazing periods (10 days each?) in one year. Not clear how long each period was/
	Methods of analysis	Means and SE of measures calculated using Restricted Maximum Likelihood (REML) to account for non- orthogonality of experiment. Sources of variance calculated using Generalised Linear Model (GLM). Regression analyses of factors affecting grazing time in the small patches. Effects of treatment, species and period analysed at plot level, using adjusted mean utilisation values to account for different sampling effort in zones. Effects of zone and interactions with zone used unadjusted data for grazed shoots. Similar techniques used as above. Because species and treatment effects estimated in both the plot and period strata there were insufficient degrees of freedom to fully

		estimate their effects. Results are quoted, but with lower confidence. Modelling of heather utilisation in relation to patch size and number.
Results		Sheep grazed for longer than deer, both species spending more time grazing in the late afternoon and evening than earlier in the day. Typical group size was greater in sheep than deer. Sheep were seen in smaller groups when the grass patch number increased. In deer group size was unaffected by vegetation pattern. Selection of grass over heather during daytime grazing was strong, but the size and distribution of grass patches significantly affected the grazing time spent on heather by sheep but not deer. In plots with one large grass patch sheep foraged on heather for only 9% of the grazing time (compared with 40% for deer), but in plots with 12 small grass patches the proportions of grazing time spent on heather were 43% for sheep and 48% for deer. On the heather the densities of grazing sheep and deer were higher in a zone of up to 5m from the edge of the grass patches than further away, and densities were higher at the edge of large patches than at the edge of small patches. This may damage heather at the edge of patches, leading to a spread of grass. In the small patch plots sheep grazing density was more evenly spread through the heather zones, suggesting use of more distant heather as sheep move between patches. Both sheep and deer showed a preference for patches with a lower proportion of dead vegetation, but sward height seemed to had little effect on patch choice for either species. Both sheep and deer showed a preference for feeding in patches that had been grazed by either species in the previous period. In the stocking rate experiment, increasing the number of sheep in a plot did not alter their feeding preferences. Group sizes at the higher stocking rate were lower in fragmented grass than in the large patch plot, and were not significantly different from the lower stocking rate in these treatments, suggesting that animals will distribute themselves to maintain their level of grass utilisation. A simple ratio of grass: heather in a moorland may not be a good predictor of heather utilisation because increased fragmentation of the available grass encourages grazers,

		particularly sheep, to graze the heather more, which may result in damaging levels of utilisation over a wider area. Heather utilisation rates varied seasonally, and increased in autumn under sheep grazing. At the plot level utilisation levels in a period were related to stocking rates, and didn't exceed 5% for a 10-day period. There was no significant interaction with other factors such as species or size and number of grass patches. Analysis at the zone level showed a significant effect with heather utilisation significantly higher in 0-5m zones round grass patches, particularly in plots with one large patch, where patch-edge is shortest. With many small patches, heather utilisation is more uniformly spread. This means damage may occur over a wider area if stocking rates are high enough to exceed damaging thresholds of utilisation. More dead shoots were also recorded in the 0-5m zone, suggesting grass patches could be extended at high sticking rates. Difference in utilisation between patch-size treatments and zones did not exactly match the patterns in location of grazing animals recorded in daylight hours – i.e. utilisation in the 5-30m zone did not increase with more frequent small patches. Heather utilisation by sheep increased as grass sward heights declined, an effected not observed with deer. The findings suggest that grazing tends to be concentrated near grass patches, particularly problematic if grass is concentrated in few large patches. Deer appear to range more freely so likely to have lower impact for similar stocking rates. Size and distribution of grass patches in dwarf shrub heath influences grazing pressure. Sheep graze in smaller groups and spend more time in heather where there are many small grass patches compared to few large patches. Sheep density on heather is highest in a zone of up to 5m around grass, and heather utilisation was found to be highest here. Increased fragmentation of grass can therefore lead to increased grazing on heather, and potentially damaging utilisation rates at high sheep numbers, over a wider area than the same area of grass concentrated in fewer large patches.
Notes	Limitations identified by author	Lack of replication and non-orthogonal experimental design. Grazing only monitored during daylight hours (16-19 hours per day), with evidence that there was a shift from grass to heather use during the night. Some heather growth between experiments 1

		and 2 in 1992 affected the assumed starting level of utilisation at expt 2.
	Limitations identified by review team	Limitations in extrapolating findings from small plot experiments to grazing behaviour of sheep on open hill, where choice of vegetation type is greater and other factors such as shelter may influence grazing choice.
	Evidence gaps and/pr recommendations for further research	Night-time heather utilisation and timing of change from grass may be different for sheep and deer, leading to greater differences in utilisation between the two species than observed here. Work needed to assess winter grazing behaviour and impact, as standing crops of grass and heather become more depleted.
	Sources of funding	Agriculture and Food research Council and Natural Environment Research Council under the Joint Agric and Environment Programme.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Gareth D Clay, Fred Worrall, Emma Clark, Evan D G Fraser
	Year	2009
	Aim of study	Hydrological responses to managed burning and grazing in an upland blanket bog
	Study design	1
	Quality score	=QA5.1 The baseline was identical for all plots – Area burnt right across in 1954 Data normalised to minimise the effect of differences due to variations in conditions on different sampling days by using the grazed/unburnt plots as a control. Depth to water table measured using cane and tape measure at least monthly until Feb 2007, when the dipwells were removed for burning on the 10year plots and returned to the same plots immediately afterwards.
	External validity	=QA5.2 Sampling continued at least monthly until Jan 2008. The study considered 33 months of data with at least 1 year before and after a burn. A total of 59 sampling visits.
Population and setting	Source population	Trout beck catchment within Moorhouse NNR. Above 500m. Geology described in detail. Mean temperatures and rain/snowfall detailed. Veg. Dominated by Eriophorum, Calluna vulgaris and sphagnum spp. Grazed by sheep at 0.6-1sheep/ha, summer months only. No burning since 1954

	Eligible population	Blanket bog
	Inclusion and exclusion criteria	
	Setting	'The Trout Beck catchment is an 11.4 sq km blanket peat area in the headwater of the River Tees.'
Methods of allocation to intervention/control	Methods of allocation	4 blocks heather moorland, each split in 6, of which 3 were enclosed to prevent grazing and 3 left unfenced. Within these blocks of 3, 3 burning regimes were randomly assigned.
	Intervention description	All blocks burnt in 1954, then 3 regimes set up: no further burning; burnt every 10 years, burnt every 20 years. The 10 year burn rotation plots were due to be burnt spring 2006, so times to examine the effect of burning and grazing at the end of the 10 year burn cycle.
	Control/comparison description	
	Sample sizes	4 blocks heather moorland, each split in 6, of which 3 were enclosed to prevent grazing and 3 left unfenced. Within these blocks of 3, 3 burning regimes were randomly assigned. In each plot, 3 dipwells inserted at least 90cm, with regular openings along the entire length.
	Baseline comparisons	All first burnt together in 1954. Normalisation by considering the grazed/unburnt plots as a control
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Depth to water table and run-off in relation to grazing/non-grazing, 10year burn and 20year burn.
	Secondary outcome measures	Continuation of Worrall (2008)
	Follow-up periods	

	Methods of analysis	Used MINITABv13 software package analysis of variance
Results		1. Shallowest water tables found on 20year burn/grazed sites. Deepest on sites never burnt. 2. In the year after a burn, water tables on that site were significantly shallower than before. 3. Hydraulic conductivity, as determined by dipwell slug tests, was significantly lower on 20year burn plots. 4. Run-off occurrence was recorded and occurred at a significantly greater frequency on sites that had recently been burnt. 'This paper demonstrates how the use of managed burning in upland settings can affect various hydrological responses of the peatland. These variations in hydrological response will have important consequences in DOC export through changes in water table and the partitioning of precipitation into runoff.'
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NE, RELU (DEFRA & SEERA)

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Burning_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Hydrological responses to managed burning and grazing in an upland blanket bog Gareth D Clay, Fred Worrall, Emma Clark, Evan D G Fraser Journal of Hydrology, 376 (2009) pp 486-495
Study Design Category	1
Assessed by & when	Alison Hiles 27/2/2013

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 ☐NA	Comments: Trout beck catchment within Moorhouse NNR. Above 500m. Geology described in detail. Mean temperatures and rain/snowfall detailed. Veg. Dominated by Eriophorum, Calluna vulgaris and sphagnum spp. Grazed by sheep at 0.6-1sheep/ha, summer months only. No burning since 1954
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 ☐NA	Comments: Characteristic of North Pennines
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 ☐NA	Comments: Characteristic of North Pennines

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 ☐NA	Comments: 4 blocks heather moorland, each split in 6, of which 3 were enclosed to prevent grazing and 3 left unfenced. Within these blocks of 3, 3 burning regimes were randomly assigned. The same sample blocks are used as Worrall (2008)
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Soil water accessed via a series of dipwells, starting April 2005, initially including no-burning and 20 year rotation plots for grazed and ungrazed plots in June. In each plot, 3 dipwells inserted at least 90cm, with regular openings along the entire length. Each opening was 4 holes at ninety degrees to each other. Care taken to avoid peat compression and sampling not done for at least 2 weeks to let peat adjust. Depth to water table measured using cane and tape measure at least monthly until Feb 2007, when the dipwells were removed for burning on the 10year plots and returned to the same plots immediately afterwards. Crestfall run-off traps were installed in Oct 2006 to intercept surface flow across the plots.
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)	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sampling continued at least monthly until Jan 2008. The study considered 33 months of data with at least 1 year before and after a burn. A total of 59 sampling visits.
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 ☐NA	Comments:
2.5 Were any other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional	✔ ☐++ ☐+ ☐-	Comments:

interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR ☐NA	
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: North Pennines, typical of upland grouse moors
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

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?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Depth to water table measured using cane and tape measure at least monthly until Feb 2007, when the dipwells were removed for burning on the 10year plots and returned to the same plots immediately afterwards.
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)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	✔ ☐++ ☐+ ☐-	Comments:

	☐NR ☐NA	
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)?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

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 ☐NA	Comments: The baseline was identical for all plots – Area burnt right across in 1954
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.	☐++ ☐+ ☐-	Comments:

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	✔ ☐NR ☐NA	
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data normalised to minimise the effect of differences due to variations in conditions on different sampling days by using the grazed/unburnt plots as a control. All pre-burn data analysed (extended dataset to Worrall (2007) The effects before and after burn investigated on the 10year plots Total dataset analysed but limited as the 10 year plots burnt part way through the study and no 10year controls were left unburnt at that point. Runoff was assessed using a kai-squared test.
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 ☐NA	Comments:
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: The baseline was identical for all plots – Area burnt right across in 1954 Data normalised to minimise the effect of differences due to variations in conditions on different sampling days by using the grazed/unburnt plots as a control. Depth to water table measured using cane and tape measure at least monthly until Feb 2007, when the dipwells were removed for burning on the 10year plots and returned to the same plots immediately afterwards. Same reservations as Worral – short term study, not covering whole burning cycle, possible confounding

		across plot boundaries as water table is continuous
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: Sampling continued at least monthly until Jan 2008. The study considered 33 months of data with at least 1 year before and after a burn. A total of 59 sampling visits.

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. Effect of grazing on delivery of moorland biodiversity

Study details	Authors	Cole et al.
	Year	2010
	Aim of study	To identify the primary habitat characteristics influencing invertebrates in year-round and summer-only sheep grazing systems; and to examine habitat-invertebrate interactions at a range of spatial scales
	Study design	Quantitative observational
	Quality score	+
	External validity	+
Population and setting	Source population	Scottish uplands - P. aquilinium/agrostis-festuca grassland and agrostis-festuca/nardus grassland
	Eligible population	30 locations chosen to represent a range of variables, including habitat composition
	Inclusion and exclusion criteria	Large-scale experiment established by SAC to manipulate grazing regimes
	Setting	Scotland (Sourhope research station)

Methods of allocation to intervention/control	Methods of allocation	Measurements from SAC study site
	Intervention description	Year-round and summer-only sheep grazing
	Control/comparison description	Comparison between year-round and summer grazing plots
	Sample sizes	9 pitfall traps at 30 sampling locations (vegetation data collected at each location, including 25 or 50 sward heights measured from each of the main patches)
	Baseline comparisons	N/A
	Study sufficiently powered	+
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Invertebrates abundance/type/size distributions, sward height
	Secondary outcome measures	30 continuous habitat variables (17 analysed)
	Follow-up periods	Grazing manipulation started in 2002, pitfall trap data collected in 2004
	Methods of analysis	Canonical Correspondence Analysis
Results		At lower spatial scales (<1m), the area of fine and broad-leaved grasses had a strong impact on mobile arthropod assemblage, at larger spatial scales (>5m), the grazing regime of the plot became more important. Sites grazed year-round had a higher relative abundance of smaller invertebrates (e.g. small predatory beetle larvae (<10mm) and small carabids (<9mm), summer only grazed sites contained larger predatory beetle larvae (>30mm), wolf-spiders, harvestmen and

		larger carabids (>15mm). At low spatial scales (<1m), vegetation height was the primary factor driving immobile invertebrate assemblage structure, at wider scales (>3m), influence of grazing regime and area of fine and broad-leaved grasses became more important. Sites with year-round grazing were associated with earthworms, leather jackets and large Limacidae slugs (>15mm), summer-grazed sites were associated with lepidopteran larvae, symphytan larvae and small Limicidae slugs and Arionidae slugs. For carabids, vegetation height was important at spatial scales of <1m, vegetation heterogeneity was significant at areas of >3m
Notes	Limitations identified by author	Care must be taken interpreting influences of grazing regime, as less obvious underlying differences between plots may have influenced invertebrate assemblages Further analysis needed to disentangle influences of plot from grazing regime Grazing pressure also influenced by altitude
	Limitations identified by review team	N/A
	Evidence gaps and/pr recommendations for further research	Replication across other sites/locations to validate data
	Sources of funding	SAC received financial support from Scottish Government Rural and Environmental Research and Analysis Directorate

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

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	
Study Citation	Cole et al. (2010)
Study Design Category	
Assessed by & when	SUSANNA PHILLIPS 25/10/2012

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 ☐NA	Comments: 2 adjacent upland plots (>40ha each) Sourhope research station (Grid ref NT8421) Plot 1 P. aquilinium on agrostis-festuca grassland Plot 2 Agrostis festuca and N. Stricta grassland
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 ☐NA	Comments: 30 locations chosen to represent range of variables (incl habitat composition)
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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 9 pitfall traps used at each of 30 locations How pitfall traps located is not described How samples selected for sward height measurements is not described

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Measurements taken from SAC study site with year-round sheep grazing and summer–only sheep grazing
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Invertebrate assemblage structure – 30 continuous habitat variables and 1 categorical variable (grazing regime)
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 ☐NA	Comments: Not reported in paper
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 ☐NA	Comments:
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Scotland-based study

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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Some subjective grouping of species by visual estimate of size Subjective assessment of vegetation type 25-50 sward height measurements – subjective selection of stem for measurement
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 ☐NA	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Invertebrates classified, size distributions, 30 continuous habitat variables (17 analysed)
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Direct measures Classification of invertebrates into mobile/immobile groupings – taxonomically broad groups, may have masked underlying mechanisms driving assemblage structure (But carabidae spp identified to species level
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Experiment started 2002, pitfall trap data collected May 2004

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 2002-2004, longer timescales may have affected results

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 ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Effect of grazing pressure on habitat
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 ☐NA	Comments: Canonical Correspondence Analysis Relationships between continuous variables analysed and highly correlated variables removed Sampling efficiency varies by habitat, so all analyses were relative not absolute The number of variables in analysis was restricted to reduce problems associated with multicollinearity
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 ☐NA	Comments: p-values given

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: Grazing regime only implemented at plot scale, therefore, influences of grazing regime and plot were confounded Comments R Pakeman: 1.2. Only two plots so there is no estimate of error as there is no replication. The locations could b e argues as representing psuedoreplication. 2.4 .As there were only two plots then counfounding factors could be plentiful. 3.6 If the species are reacting to structure then the follow up time is perhaps meaningful. However, those reacting to vegetation change are unlikely to be affected over such a short period. 4.1. Power is zero as no replication of main plot treatment. 4.3. The methods fail to mention whether the permutation tests took into account the design of the experiment - i.e. the need to pemute within main plots. 5.1 and 5.2 I disagree with the overall gradings on this. I would score a'-'' Ultimately scored 2+
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: Generalisable to similar grassland habitats

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ___Uplands_____

Name of Review Sub-topic (if any): ____Moorland Grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Common, T. G., Wright, I. A. & Grant, S. A. (1998). The effect of grazing by cattle on animal performance and floristic composition in Nardus-dominated swards. Grass and Forage Science, 53, 260-269
Study Design Category	2
Assessed by & when	D Martin 7/12/12

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: Nardus- dominated grassland on rough hill grazings. Previous grazing experiments and results summarised
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: U5 grassland. Likely to be typical of the wider habitat, but choice of study areas limited by practical considerations and in this case is a research station.
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: Likely to be representative of the area, but selection probably subjective

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: Two treatments with two replicates. Not sta6ted as random, but less relevant for only two replicates
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Two treatments were implemented based on target sward heights. Grazed with spring-calving blue-grey cattle
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: Treatments in place for 5 years
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:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Location is Cheviots on Scottish Borders. Likey to be fairly representative of species-poor rough hill grazing, but
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Grazing of lactating cattle on hill land is not now common practice. Especially at the higher rate. The agronomic implications of this are part of the aims of the experiment.

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: Measurements were objective. Sward heights measured using sward sticks from forty points per plot. Floristic composition from inclined point quadrats on transects (restricted random). Measures of Nardus utilisation (proportion of utilisation) and closeness of grazing (lamina length) at randomly chosen points. Cow liveweight measurements made and dietary measurements of org matter intake and digestibility and diet floristic composition from samples obtained from fistulated cows in 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:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Appears so from the Measurements section.
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: All treatments in place for 5 year period
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments:

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: Percentage cover of Nardus, broad and fine-leaved grasses were shown to be similar in the different treatments at the start of the experiment.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?		Comments:

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	
4.3 Were the estimates of effect size given or calculable?	☐NR	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: Animal data analysed using residual maximum likelihood fixed effects models. Change in floristic composition, utilisation and grazing effects on N, P and K in the leaves of Nardus and A cap analysed using ANOVA.
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 ☐NA	Comments:
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: Well designed study, but based only two treatment replicates.
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:

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Common, T. G., Wright, I. A. & Grant, S. A.
	Year	1998
	Aim of study	To explore levels of *Nardus* utilisation by lactating cattle and resulting effects on animal performance and floristic composition.
	Study design	2
	Quality score	+
	External validity	++
Population and setting	Source population	Nardus- dominated grassland on rough hill grazings. Previous grazing experiments and results summarised
	Eligible population	U5 grassland. Likely to be typical of the wider habitat, but choice of study areas limited by practical considerations and in this case is a research station.
	Inclusion and exclusion criteria	NA

	Setting	Sourhope research station, Cheviot Hills, Scotland. Altitude of 520m.
Methods of allocation to intervention/control	Methods of allocation	Two treatments with two replicates. Not stated as random, but less relevant for only two replicates
	Intervention description	Two treatments were implemented based on target sward heights (4-5cm and 6-7cm). Grazed with spring-calving blue-grey cattle
	Control/comparison description	No control as such, just comparison of two treatments.
	Sample sizes	Treatment areas 5.1 and 7.15 ha for short and tall treatments respectively. Twice weekly sward heights from 40 points per plot. Veg composition from at least 25 point contacts at 24 locations per plot. Utilization measured on five tillers from 40 locations per plot, and lamina length at 80 randomly chosen leaves. Dietary measurements three times per year. Nardus tussocks measured in 16 2mx2m quadrats per plot at end of experiment.
	Baseline comparisons	Measurements made in first year – floristic composition shown to be similar at start.
	Study sufficiently powered	Not reported
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Levels of Nardus utilisation and diet composition of animals. Florisitc composition and change over time. Animal performance.
	Secondary outcome measures	Nardus tussock size, N, P and K concentrations of Agrostis and Nardus leaves.
	Follow-up periods	Experiment in place for 5 years.

	Methods of analysis	Animal data analysed using residual maximum likelihood fixed effects models. Change in floristic composition, utilisation and grazing effects on N, P and K in the leaves of Nardus and A cap analysed using ANOVA.
Results		Mean annual stocking rates required to maintain the shorter inter-tussock height (4-5cm) declined from 1.8 to 1.18 per ha, and the tall(6-7cm) varied between 0.76 and 1.04 per hectare. Herbage intake in the short treatment was 0.65 of that of cows grazing the tall treatment. Cows on the tall treatment consistently produced more milk, and had an overall increase in liveweight compares with a small loss on the short treatment. The cover of Nardus and broad-leaved grasses declined significantly during the experiment, and was greater on the short treatment. Cover of fine-leaved grasses declined on the short treatment but not on the tall. Other groups of mosses, sedges and herbs did not change significantly, other than *Molina* which was present in small patches and quickly eliminated from the sward. Cows grazed a higher proportion of *Nardus* on the short treatment, and to a shorter height. More tillers were grazed in July than at the end of the season. The percentage frequency of live Nardus was greater and dead material less than in the sward, and these differences increased over time in each treatment. Leaf concentrations of N and P were seen to reduce as a result of grazing in *Agrostis*, to a greater extent than in *Nardus*, with no significant difference between treatments. There were more tussocks with an area of less than 140cm^2 in the short treatment, with the difference between treatments greatest at the smallest tussock sizes. More *Nardus* is ingested at the low sward treatment, confirming previous findings that this species is less preferred than other grasses. The increase in the proportion of *Nardus* in diet over time was associated with a decrease in dead material in the diet. The sward initially contained a high proportion of ungrazed tussocks as a result of sheep grazing. As dead material was removed, the proportion of live material selected increased. The decline in more palatable between-tussock grasses, in contrast to other

		studies, may be due to inherently low fertility or altitude and exposure of the site. Five years of cattle and calves grazing resulted in decline in *Nardus* cover by almost half, and decrease in tussock size especially with the heavier grazing treatment, based a target sward height of 4-5 cm. Cows on both treatments ingested a greater proportion of *Nardus* than was present in the sward, and the proportion of live to dead material in diet increased over time as the sward adjusted from previous sheep grazing. The heavy grazing regime was effective in controlling *Nardus*, but not compatible with animal performance. The moderate grazing treatment (6-7 cm) may allow acceptable animal performance, and some control of *Nardus*. Increases in inter-tussock palatable grasses was not observed in this study, which may be an effect of low soil fertility.
Notes	Limitations identified by author	
	Limitations identified by review team	Only two replicates. No control (typical sheep grazing?) although starting point reflects history of sheep grazing.
	Evidence gaps and/pr recommendations for further research	Research on the interactions between species composition, levels of defoliation, climatic condition and nutrient supply from the soil
	Sources of funding	Scottish Office Agriculture, Environmental and Fisheries Department

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Cooper, A., McCann, T. & Power, J
	Year	1997
	Aim of study	To develop a regional model of distribution, composition and management of heath and mire vegetation in the Northern Ireland uplands, based on multivariate land classification and a structured land cover and vegetation sampling programme.
	Study design	2
	Quality score	++
	External validity	+
Population and setting	Source population	Upland land classification squares identified through the Northern Ireland Countryside Survey. Based on NCC habitat classification. Not described in detail but covers the main upland land cover types.
	Eligible population	Six upland study areas identified for their area of statutory designations. Will cover main upland land cover types and vegetation groups.
	Inclusion and exclusion	Comments: The six study areas sampled via a stratified random sample approach of

	criteria	land cover classes within the areas. Field study involved mapping the extent of heath and mire communities according to standard definitions.
	Setting	Northern Irish uplands
Methods of allocation to intervention/control	Methods of allocation	Correlative study/ survey approach
	Intervention description	Various management variables are assessed on a sample basis, including grazing assessed as H, M, L
	Control/comparison description	N/A
	Sample sizes	Land cover sampling based on random sample of 628 25 ha grid squares. Species data and env and management variables recorded in 643 nested $4m^2$ and $200m^2$ quadrats.
	Baseline comparisons	N/A
	Study sufficiently powered	N/A, but large sample selected
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Land cover type and vegetation composition
	Secondary outcome measures	N/A
	Follow-up periods	N/A
	Methods of analysis	Twinspan of quadrat data and comparison of end-groups to NVC. CCA of mire quadrats, with stepwise forward selection carried out on a range of environmental, geographic and management variables to determine which explained greatest variation. Non-

		significant variables removed from the regression analysis.
Results		The most important environmental variables accounting for variation in heath and mire communities were soil waterlogging followed by grazing intensity and slope. M17 (*Trichophorum cespitosum-Eriophorum vaginatum* blanket mire) and M19 (*Calluna vulgaris-Eriophorum vaginatum* blanket mire) emerged as vegetation of deeper, lightly grazed peats. Species associated with highest grazing intensity included *Sphagnum auriculatum, Nardus stricta, Carex panacea* and *Drosera rotundifolia.* Light grazing favoured *Calluna vulgaris, S subnitens, S capillifolium, Eriophorum vaginatum* and *Empetrum nigrum.* Regional variation in vegetation characteristics of peatland could be linked to differences in the main variables of grazing intensity, peat wetness and slope, and secondary variable of peat depth. Some variables that may be important in terms of local condition (erosion, peat-cutting and drainage ditches) were not significant in explaining regional differences. Land cover and vegetation data was used to asses variation in upland heath and bog communities in Northern Ireland. Grazing was shown to be the main management factor associated with variation in species composition of between different upland areas, along with peat wetness and slope. Light grazing favoured dwarf shrub, hare's-tail cotton grass and some Sphagnum species, where heavier grazing favoured mat grass but also carnation sedge, round-leaved sundew and S auricualtum. Blanket bog communities (*Trichophorum cespitosum-Eriophorum vaginatum*) and M19 (*Calluna vulgaris-Eriophorum vaginatum*) were associated with lightly grazed, deeper peats.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for	Future peatland management strategies need to consider spatial variation at the landscape scale, including considering the association between types of management

	further research	and site location attributes.
	Sources of funding	

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

Name of Evidence Review: ___Upland_____

Name of Review Sub-topic (if any): ___Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Cooper, A., McCann, T. & Power, J. (1997). Regional variation in the cover, species composition and management of blanket bog. Landscape and Urban Planning 37, 19-28
Study Design Category	2
Assessed by & when	D Martin

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: Upland land classification squares identified through the Northern Ireland Countryside Survey. Based on NCC habitat classification. Not described in detail but covers the main upland land cover types.
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: Six upland study areas identified for their area of statutory designations. Will cover main upland land cover types and vegetation groups.
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: The six study areas sampled via a stratified random sample approach of land cover classes within the areas. Field study involved mapping the extent of heath and mire communities according to standard definitions.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments: Correlative study/ survey approach
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Environment and site management data recorded. Not clear to what detail for many factors, or just presence/ absence. In general appropriate for this strategic-level study. Grazing intensity recorded as H, M, L.
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?	☐NA	Comments:
2.5 Is the setting applicable to the UK?	☐+	Comments: NI upland areas. The range of upland habitats present are very similar to rest of UK

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: Estimates of wetland habitats from analysis of Land cover data. Vegetation sampled in 643 nested quadrats within heath and mire land cover types. Number of quadrats were proportional to the area of each land cover type.
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: In relation to the objectives of documenting distribution and composition of heath

Were all important positive and negative effects assessed?		and mire habitat.
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?	☐NA	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?	☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++	Comments:
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: Twinspan of quadrat data and comparison of end-groups to NVC. CCA of mire quadrats, with stepwise forward selection carried out on a range of environmental, geographic and management variables to determine which explained greatest variation. Non-significant variables removed from the regression analysis.
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: p values for testing of significance of variables in the model
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐++	Comments: Not experimental study, but based on large stratified random sample of vegetation, and well designed strategic land classification programme.

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: Habitats are typical, but some management and environmental factors may differ between NI and rest of UK

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. effect of grazing on biodiversity and ecosystem services g. effect of different types of livestock

Study details	Authors	Critchley et al
	Year	2008
	Aim of study	To assess the effect of two sheep only and two mixed (cattle and sheep) grazing regimes on vegetation and livestock performance when applied to heterogeneous degraded wet heath
	Study design	Quantitative experimental
	Quality score	-
	External validity	-
Population and setting	Source population	Wet heath
	Eligible population	ADAS Redesdale, mainly M15 communities
	Inclusion and exclusion criteria	Degraded wet heath communities
	Setting	Northumberland

Methods of allocation to intervention/control	Methods of allocation	Not reported
	Intervention description	4 grazing treatments: 1. Low sheep (0.66 sheep/ha minus 25% oct-feb inclusive) - LS 2. Low sheep plus cows (as 1. Plus 0.75 cows summer only)- LSC 3. High sheep (1.5 sheep/ha minus 25% oct-feb inclusive) - HS 4. High sheep plus cows (as 3. Plus 0.75 cows summer only) - HSC Sheep regimes from 1995, cows from 2003. Length of summer cattle grazing dependent on when calluna started to be grazed varied from 4 weeks to 9-10 weeks.
	Control/comparison description	Comparison between four treatments
	Sample sizes	Vegetation sampling - Total of 196 quadrats for vegetation sampling Livestock data - 18-38 sheep per plot and 16-22 cattle per cattle grazed plot
	Baseline comparisons	Baseline data collected in 2003
	Study sufficiently powered	No replication
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Percentage top cover, grazing indices, sward height, livestock liveweight
	Secondary outcome measures	N/A

	Follow-up periods	Sheep grazing from 1995, cows from 2003, data collected 2003-2006
	Methods of analysis	Detrended Correspondence Analysis used for change in species cover data. Effect of time also tested by applying partial Redundancy Analysis, using time as the explanatory variable and quadrats as covariables Mean vegetation height and differences in vegetation type over time were analysed using multivariate repeated measures ANOVA Livestock data analysed using REML
Results		Molinia declined significantly in 1st year after introduction of cows (where dominant or co-dominant in sward only) Molinia increased in both sheep only paddocks (no p-values given). No evidence of increase in calluna The paddocks with cows showed a trend for calluna type vegetation to move towards calluna/molinia (LSC $p<0.01$ and HSC $p<0.05$) and molinia type to move towards c. Nigra and n. Stricta (LSC and HSC $p<0.01$). In the HS calluna/molinia vegetation and n. Stricta vegetation moved towards molinia ($p<0.01$). Little change recorded in LS. In mixed paddocks, the grazing index was higher in molinia type vegetation than in calluna type (LSC $p<0.001$; HSC $p<0.01$), no corresponding difference was found in sheep only paddocks Cattle in the HSC paddock had lower daily liveweight change than those in the LSC (no p-value reported). Ewe condition score did not differ significantly between paddocks, although weights were lower in high stocking rate paddocks. No significant paddock effect on lamb birth weight or lambing percentage, but weaning weights significantly higher in lower stocking rate paddocks
Notes	Limitations identified by author	Calluna covers in paddocks may have been affected by previous sheep grazing regimes Restoration by grazing alone needs longer timescales than studied in this research

		Not replicated therefore effect of stocking not explicit
	Limitations identified by review team	A number of explanatory variables were not investigated – particularly the effect of weather variables
	Evidence gaps and/pr recommendations for further research	Validation through replication on other wet heath sites Application of study to other habitat types
	Sources of funding	DEFRA, CCW and Natural England

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	a. effect of grazing on biodiversity and ecosystem services g. effect of different types of livestock
Study Citation	Critchley et al (2008)
Study Design Category	Quantitative experimental
Assessed by & when	SUSANNA PHILLIPS 08/11/12

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 ☐NA	Comments: Wet heath Soil types described 260-350m AOD
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 ☐NA	Comments: ADAS Redesdale, Northumberland – grid reference given Mainly M15, small scale variation in relative abundance of species described
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 ☐NA	Comments: Points not randomly selected – a rectangular grid of 196 points at 75m spacing was used with data recorded from 1x1m quadrat at each sample point Inclusion criteria - degraded habitat

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 ☐NA	Comments: "four paddocks each with a different stocking regime" – method of allocation not recorded Spatial grazing patterns related to slope accounted for by upslope-downslope layout of paddocks Bias between habitat type minimised by adjusting fencelines
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Management interventions described in sufficient detail to replicate
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)	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No deviation from experimental design reported. Sheep grazing treatments applied since 1995, cows introduced in 2003. Pre-2003 stocking regimes may have affected results
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 ☐NA	Comments: No contamination reported, assumed exposure as experimental design
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 ☐NA	Comments: No additional intervention reported

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Study site located in Northumberland, applicable to other similar UK habitats
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Range of stocking densities/mix of livestock type typical of UK moorland grazing practices

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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Predominantly objective measures: Percentage top cover of vegetation estimated using sighter with cross wires in centre of each cell Grazing indices for calluna and molinia as proportion of occupied cells in which grazed shoots were present Sward height measured at 5 random locations in quadrat – not reported how selected Liveweight and tactile condition body scores (subjective measure – not reported whether this measure was validated) of sheep at mating and pregnancy scanning, lamb liveweights, cattle liveweights Fieldwork (quadrats) carried out in same order & paddocks recorded simultaneously to avoid confounding date of recording with paddock
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)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: All outcome measures reported on
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ ☐+ ☐-	Comments: Ecological impact considered only in terms of grazing indices, other herbivore impact (e.g. trampling) not assessed Minor temporal trends may have reflected annual

	☐NR ☐NA	variation in temperature and rainfall but data not presented Earlier outbreak of heather beetle may have affected trends for decreasing calluna values – effect not reported on
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)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Direct measures assessed
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data recorded at similar timings throughout year for all four treatments, however exposure times to sheep/cattle grazing treatments varied (see 2.3), and this may have distorted results
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 2003-2006, sufficient time to identify structural changes in vegetation, species composition may have needed longer timescales to identify trends

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 ☐NA	Comments: Baseline data collected in 2003, some variation in vegetation parameters, reports on changes within a plot should be valid, and the limited extent of differences at baseline may also allow comparison of magnitude of change between plots
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: Power calculation not recorded. No replication Measurements from livestock based on relatively low

standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐- ☐NR ☐NA	numbers
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Detrended Correspondence Analysis used for change in species cover data. Effect of time also tested by applying partial Redundancy Analysis, using time as the explanatory variable and quadrats as covariables Mean vegetation height and differences in vegetation type over time were analysed using multivariate repeated measures ANOVA Livestock data analysed using REML
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 ☐NA	Comments: p-values and standard errors given
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: Predominantly objective measures, although a number of explanatory variables not investigated. No replication. Vegetation data not taken from randomly selected sample points. Cattle grazing plots received shorter exposure times than sheep
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: '-' assigned because of low levels of internal validity of the research

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 _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Critchley, Mitchell, Rose, Griffiths, Jackson, Scott & Davies. Re-establishment of *Calluna vulgaris* in an eight-year grazing experiment on upland acid grassland. Journal for Nature Conservation, in press.
Study Design Category	2
Assessed by & when	D Martin & DATE]

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: Some description of degraded upland heathland dominated by Mat grass.
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: Site typical of degraded upland moorlands within the UK. over-grazing resulted in decline in Calluna since 1970s
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: The paddocks containing study plots were set up to be representative of the degraded heathland, grass dominated habitat, and subject to typical ESA type grazing treatments.

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: Grazing treatment allocation to paddocks was randomised and grazing regimes replicated. This was not the case at *Molinia* (Redesdale) site. (This paper reports on a continuation of work at two sites reported by Mitchell et al, 1998, but only Pwllpeiran work extended.) *Nardus* (Pwllpeiran) site (dominated by *Nardus, Agrostis, Festuca,* with some *Vaccinium*): -3 blocks of land x 3 fields (5-7ha) in each block -each block, 3 fields randomly assigned to: 'cattle'; 'mixed'; 'sheep' -each field, 6 10x10m plots in areas with similar vegetation
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: September 2002: plots randomly assigned to one of 3 disturbance treatments: 'undisturbed'; 'rotavation'; 'trampling'. March 2003: plots had 2 sub-treatments applied: -*Calluna* seed on half of each plot; -No grazing (fencing) on half of each plot. Within each quarter plot, 1 4x4m sub plot established within which all recording carried out. Within each sub-plot, 9 1x1m permanent quadrats to record *Calluna* establishment and bare ground. In 2010 three plants per plot were cut at random and morphological measurements made. Details of materials and methods given – in sufficient detail to replicate.
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: This paper reported on eight years of grazing treatment on the grazed plots, extending the work reported in Mitchell et al 2008.
2.4 Was contamination acceptably low?	☐++	Comments: Carefully controlled experiment – no

Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?		indication of contamination
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:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Typical of degraded upland grass-moorland habitats. But this study only looks at one site. Restoration of dwarf shrub was aim rather than specific NVC community. Intended that this broader aim would allow results to be applicable more widely within UK.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Stocking rates fairly typical for this type of land (e.g. 0.5 cow/ha July and August; 1 – 1.5 ewes all year round). But cattle not always available on British upland farms. Disturbance treatments: 'undisturbed'; 'rotavation'; 'trampling' are novel, and aimed at restoration so not typical agricultural work on moorland. Intended to inform other restoration projects.g

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: Cover of all species and C vulgaris recorded in 4m2 quadrats in each sub-plot and height measurements in each of 16 1m quadrats. Cover measurements judged by eye so relatively subjective. In 2010 additional morphological and dry weight measurements on Calluna. Comparison measurements made outside of plots to test whether they were subject to preferential grazing.
3.2 Were all outcome measurements		Comments:

complete? 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: All important effects assessed: -Seed-bank composition -Effects of treatments on *Calluna* establishment: heather seeding disturbance treatments, bare ground removing grazing different grazing regimes
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: Directly relevant to objectives.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: All treatments subject to same timescales.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: This work extended study to 8 years, so enough time to judge heather establishment and community development

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: Similar, and replicated and randomised treatments
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?	☐NR	Comments:
4.3 Were the estimates of effect size given or calculable?	☐NA	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: Separate analyses for each site. -Used generalized linear mixed models (GLMM) -'Fixed effect' both sites: disturbance, fencing, seeding, visit, and their interactions; *Nardus* site: block and grazing regime; *Molinia* site: field. -'Random-effect': plot -'Continuous variables': bare ground, *Calluna* morphology
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: p-values quoted throughout results section.
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: Study design appears unbiased. Potential confounder – residual seed-bank was thoroughly investigated. Acknowledged that cattle grazing (summer only) could be confounded with seasonality
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)?	☐++	Treatments carried out in small plots but were developed to be applicable at larger scales. The limiting factors are the economics of applying the initial treatment and thereafter maintaining appropriate stocking regimes. If implemented at a landscape scale, an increase in the use of cattle grazing – considerable change in farming practices in the British uplands.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Critchley, Mitchell, Rose, Griffiths, Jackson, Scott & Davies
	Year	In Press
	Aim of study	To report the longer term (8-year) effects of grazing treatments and restoration treatments on Calluna development and community composition, in terms of progress towards target plant community.
	Study design	2
	Quality score	++
	External validity	++
Population and setting	Source population	Some description of degraded upland heathland dominated by Mat grass. Site typical of degraded upland moorlands within the UK. over-grazing resulted in decline in Calluna since 1970s
	Eligible population	The paddocks containing study plots were set up to be representative of the degraded heathland, grass dominated habitat, and subject to typical ESA type grazing treatments.
	Inclusion and exclusion	This paper reports on a continuation of work at two sites reported by Mitchell et al, 1998, but

	criteria	only Pwllpeiran work extended here.
	Setting	Pwllpeiran Research Farm, Cambrian Mountains, Mid Wales
Methods of allocation to intervention/control	Methods of allocation	Grazing treatment allocation to paddocks was randomised and grazing regimes replicated. This was not the case at *Molinia* (Redesdale) site (not reported here).
	Intervention description	3 blocks of land x 3 fields (5-7ha) in each block -each block, 3 fields randomly assigned to: 'cattle'; 'mixed'; 'sheep' -each field, 6 10x10m plots in areas with similar vegetation. September 2002: plots randomly assigned to one of 3 disturbance treatments: 'undisturbed'; 'rotavation'; 'trampling'. March 2003: plots had 2 sub-treatments applied: -*Calluna* seed on half of each plot; -No grazing (fencing) on half of each plot.
	Control/comparison description	Undisturbed, unseeded, sheep grazed.
	Sample sizes	6 plots established in each of the three grazing treatments. Two replicates of each disturbance treatment per paddock, and grazed vs ungrazed and seeded vs unseeded applied in 2x2 factorial.
	Baseline comparisons	Vegetation measurements made before original treatments imposed.
	Study sufficiently	Not reported

	powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	*Calluna* cover in each sub-plot and change in *Calluna* cover over time. Height of Calluna and number of shoots, and dry weight of a shoot sample compared between grazing treatments.
	Secondary outcome measures	Similarity to target heathland community. Soft rush occurrence.
	Follow-up periods	Measurements made at three years and eight years.
	Methods of analysis	Used generalized linear mixed models (GLMM) -'Fixed effect': disturbance, fencing, seeding, visit, and their interactions; block and grazing regime. -'Random-effect': plot -'Continuous variables': bare ground, height; *Calluna* morphology Modified Bray and Curtis similarity index used to compare cover data with target community.
Results		Grazing exclusion and cattle grazing had a significant effect on heather morphology, with significantly greater height, number of shoots and dry weight than in either the sheep only or mixed grazing plots. Heather cover was highest in seeded and rotovated or trampled plots protected from grazing (25-30% cover), and next highest in the same combination of treatments where cattle only grazed (20-25%). Plots grazed by sheep only had the lowest heather cover, with seed addition having much less effect in these plots. Change in heather cover over the previous four years was significant in ungrazed and cattle only plots. In disturbed and seeded plots the similarity to target vegetation was highest in ungrazed and cattle only grazed plots. Overall it would appear that grazing exclusion for an extended period is the best option for heathland restoration, but summer cattle grazing provides a viable alternative
Notes	Limitations identified by author	Still relatively short-term in habitat restoration terms. Lack of knowledge / study on recolonsation of later colonising species necessary for closer correspondence to

		heathland communities.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Economics of grazing removal or cattle introduction. Timescales of recovery to target community still unclear. When can grazing, especially sheep grazing, be introduced?
	Sources of funding	Defra

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	g. do different types of livestock affect moorland habitats differently? (a. what is the effect of grazing on the delivery of moorland biodiversity?)

Study details	Authors	De Gabriel et al.
	Year	2011
	Aim of study	To quantify the relative effect of different herbivore species, vegetation structure and rainfall on heather utilisation, species richness and evenness (alpha diversity) and beta diversity
	Study design	Quantitative observational/correlation
	Quality score	- (changed to + after comments from R Pakeman)
	External validity	- (+)
Population and setting	Source population	Upland heather moorland
	Eligible population	Grass-heather mosaics
	Inclusion and exclusion criteria	≥50% cover of heather Sites previously subject to Rapid Habitat Impact Assessment Sheep had been removed from one site within each pair and all sites supported deer

		populations
	Setting	Scotland
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	Eight pairs of sites, sheep had been removed from one site within each pair while red deer grazing continued at all sites
	Control/comparison description	Comparison between deer-only grazed sites and sheep/deer grazed sites
	Sample sizes	40 samples per 10x10m plot 6-12 10x10m plots per site 8 pairs of sites
	Baseline comparisons	N/A
	Study sufficiently powered	+
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Heather utilisation, sward height, dwarf shrub height, covers
	Secondary outcome measures	Dung count
	Follow-up periods	Sheep removal 1960-2002, surveys carried out 2007/2009
	Methods of analysis	Linear mixed effects models

Results		More deer dung was present where sheep were absent from a site (p=0.004) and amount of deer dung was positively correlated with length of time since sheep removal at a site scale (p=0.05) At a site scale, heather utilisation was positively correlated with the amount of deer dung, percentage of grass and mean smooth grass height. Length of time since sheep removal had no effect on grazing impacts at any spatial scale. Heather was taller where sheep were present at the site scale, but no effect of herbivore was found on grass height. The amount of smooth grass was positively correlated with amount of sheep dung at a site Alpha diversity was positively correlated with percentage of grass and the amount of sheep dung Beta diversity was higher where sheep were present
Notes	Limitations identified by author	None
	Limitations identified by review team	Inter- and intra-rater reliability for subjective measures unreported Potential bias introduced by analysis of sites where graziers had chosen to remove sheep
	Evidence gaps and/pr recommendations for further research	Further research to determine how similar changes in grazing regime affect abundance & assemblage of other species (eg invertebrates) and their potential interactions with plant communities
	Sources of funding	Scottish government/James Hutton Institute

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

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	g. do different types of livestock affect moorland habitats differently? (a. what is the effect of grazing on the delivery of moorland biodiversity?)
Study Citation	DE GABRIEL ET AL (2011)
Study Design Category	
Assessed by & when	SUSANNA PHILLIPS 08/10/2012

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 ☐NA	Comments: 16 upland sites in Scotland Heather/grass mosaics
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 ☐NA	Comments: Square plots selected to overlap with Nolan et al (2002) Only sites with more than 50% heather cover selected Attempts to minimise variation in environmental variables
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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 10*10m plots selected randomly within site Bias – w-walk to select sample points Randomly thrown pin-frame 2x2m quadrat for % cover not randomly located - in sw corner of 10x10m plot

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Review of sites where sheep were already removed – potential bias Replication at 3 spatial scales
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Herbivore abundance – heather utilisation, sward height, dwarf shrub height, covers
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 ☐NA	Comments: Sheep re-introduced on one site – excluded from analysis
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 ☐NA	Comments: Altitude/precipitation accounted for, other environmental variables not considered
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Scotland Applicable to deer grazed sites

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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Subjective assessment of grazing levels – observer bias not validated Vegetation heights objective, but selection of vegetation subjective within w-walk Frequency of samples approximated to cover from 40 sample points – unclear if valid at this sample size
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 ☐NA	Comments: Data reported for all outcomes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Spp diversity, heights, calluna cover – appropriate to meet objectives of study
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Dung count used as proxy for herbivore density – appropriate for comparisons between sites Frequency used to approximate to cover
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Surveys mar-may & jun-aug (3 month sampling window – not reported if an individual site is surveyed at similar time within window each year)

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Surveys carried out 2007/2009 following sheep removal 1960-2002

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 ☐NA	Comments: Effect size given for each explanatory variable 40 samples per 10x10m plot 6-12 10x10m plots per site 8 pairs of sites
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Herbivore, habitat and environmental variables
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 ☐NA	Comments: Precipitation data interpolated for altitude Linear mixed effects models
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 ☐NA	Comments: p-values given

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: Observer bias not validated Objective measures, but samples selected subjectively Subjective measure of grazing level on calluna
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: See 5.1 R Pakeman comments: I think these have been harshly marked as the utilisation measure isn't subjective as the stems are compared to adjacent ungrazed ones. There are wide categories to speed things up, which reduces precision but does not increase bias. Changed to +

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): ____Moorland Grazing_____

Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Deléglise, C., Loucougaray, G & Alard, D. (2011). Effect of grazing exclusion on the spatial variability of subalpine plant communities: A multiscale approach. Basic and Applied Ecology 12, 609-619.
Study Design Category	2
Assessed by & when	D Martin 12/11/12

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: Subalpine areas of France. Not described in 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: Likely to be fairly representative of at least part of the range of sub-alpine communities – calcareous grassland, mesic grassland and heath grassland. Described in some detail in terms of characteristic and dominant species.
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: Study used pre-existing ungrazed exclosures with paired grazed plots, three pairs per vegetation type. Pairs chosen to be similar in pre-exclosure grazing history, environmental conditions and species composition and diversity. Initial site selection likely to have been subjective.

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: Not described whether treatment was randomised within pairs. Three replicates per vegetation type
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐-	Comments: No details given of the grazing treatments at each plot or site. A generalisation is made that most areas are now moderately summer grazed, removing c50% of above ground net primary production.
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: Treatments in place for 20 years. Not clear whether grazing was constant over this period.
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: None reported. There is the possibility that some grazed plots were not grazed in every year, but given length of the experiment this is unlikely to be significant unless it was for a number of years.
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: Study outside of UK and sites above the maximum UK altitude, so not representative. There are however analogous vegetation types in UK, but with sometimes different but related dominants.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Summer sheep and cattle grazing will be similar to some UK grazing units, although in many areas of UK hill sheep are grazed for most of the year.

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: Ground cover of species measured in random quadrats. Cover estimates were subjective, and adjusted up or down to add up to 100% in each quadrat. Leaf traits measures were objective.
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: All measurements made in one year
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Structural measures?
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: Yes, treatments in place for 20 years (any variation in grazing regimes over this period not reported, but assumed to be fairly consistent). Enough time for grazing-related differences to be detected.

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: Reported as similar when plots were set up, but no data presented.

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?	☐NR	Comments: Good sample size – three replicates from each vegetation type. Between 8 and 40 quadrats from large to small scale respectively
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments:
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: Quadrat species data used to calculate Euclidean distance, averaged among all pair-wise comparisons for each grain size. Aggregated leaf-trait values calculated for each quadrat, weighted by relative abundance of species, and spatial variability calculated at each grain size. These variables tested against grain size for grazed and ungrazed in each grassland. Effects of plant community and grazing treatment tested using mixed effects ANOVA.
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: p values given for ANOVAs and regressions
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: Well designed and replicated, with adequate sampling. Details of grazing treatment and variation in it (over time and between sites) not given
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: Principles in terms of effects of abandonment are generalisable to similar grassland and grass-heath communities. Setting not UK, but analogous communites

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): ____Moorland Grazing_____

Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Deléglise, C., Loucougaray, G & Alard, D. (2011b). Spatial patterns of species and plant traits in response to 20 years of grazing exclusion in subalpine grassland communities. Journal of Vegetation Science 22, 402-413
Study Design Category	2
Assessed by & when	D Martin 16/11/12

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: Subalpine areas of France. Not described in 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: Likely to be fairly representative of at least part of the range of sub-alpine communities – xeric calcareous grassland, mesic grassland and heath grassland. Described in some detail in terms of characteristic and dominant species (table 1 on paper).
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: Study used pre-existing ungrazed exclosures with paired grazed plots, three pairs per vegetation type. Pairs chosen to be similar in pre-exclosure grazing history, environmental conditions and species composition and diversity. Initial site selection likely to have been subjective.

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: Not described whether treatment was randomised within pairs. Pairs are 500m to several km apart. Three replicates per vegetation type
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐-	Comments: No details given of the grazing treatments at each plot or site. A generalisation is made that most areas are now moderately summer grazed, removing c50% of above ground net primary production.
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: Treatments in place for 20 years. Not clear whether grazing was constant over this period.
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: None reported. There is the possibility that some grazed plots were not grazed in every year, but given length of the experiment this is unlikely to be significant unless it was for a number of years.
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: Study outside of UK and sites above the maximum UK altitude, so not representative. There are however analogous vegetation types in UK, but with sometimes different but related dominants.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Extensive to moderate summer sheep and cattle grazing will be similar to some UK grazing units, although in many areas of UK hill sheep are grazed for most of the year.

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: Canopy height of dominant vegetation measures in 100 contiguous 10cm quadrats from each of two perpendicular transects crossing the plot. Adequate to measure the required spatial pattern? Four plant traits measured with 20 or 10 replicates per species, for all species necessary to reach 80% cover in 1m 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: All measurements made in one year
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: Do belt transects adequately pick up patch size and spatial pattern?
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: Yes, treatments in place for 20 years (any variation in grazing regimes over this period not reported, but assumed to be fairly consistent). Enough time for grazing-related differences to be detected.

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: Reported as similar when plots were set up, but no data presented.

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: States that the sampling design allowed enough pairs of observations (for calculating correlation between variables at different distance classes) to ensure the power of the test for all distance classes up to 500cm.
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments:
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: Cummulative plot scale abundance of all species belonging to five life forms calculated. Aggregated leaf-trait values calculated for each quadrat, weighted by relative abundance of species. Measure of autocorrelation (Moran's I) calculated to test the degree of correlation between the values of a variable as a function of spatial location. Calculated for different distance classes up to 500cm (see 4.2) Derived measures of Grain and contrast between patches. Effects of treatment, plant community and interactions tested using mixed-effects ANOVA.
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: p values given for ANOVA.
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: Well designed and replicated, with adequate sampling. Details of grazing treatment and variation in it (over time and between sites) not given. Possible confounding of livestock type – cattle grazing with sheep at heath site.
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: Principles in terms of effects of abandonment are generalisable to similar grassland and grass-heath communities. Setting not UK, but analogous communities

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Deléglise, C., Loucougaray, G & Alard, D
	Year	2011
	Aim of study	To compare species diversity (richness and evenness) and vegetation spatial heterogeneity of subalpine grassland communities between traditionally grazed plots and their long-term (>20 years) ungrazed equivalents by sampling vegetation at seven spatial scales.
	Study design	2
	Quality score	+
	External validity	++
Population and setting	Source population	Subalpine areas of France. Not described in detail
	Eligible population	Likely to be fairly representative of at least part of the range of sub-alpine communities – calcareous grassland, mesic grassland and heath grassland. Described in some detail in terms of characteristic and dominant species.

	Inclusion and exclusion criteria	Experiments based on long-term plots, with original selection not reported. Pairs chosen to be similar in pre-exclosure grazing history, environmental conditions and species composition and diversity.
	Setting	Grassland habitats in Hauts-Plateaux du Vercors Nature Reserve, and heath at Alp d'Huez ski resort, Western Alps, France. Altitude 1600-1800 m asl. High precipitation and snow cover duration.
Methods of allocation to intervention/control	Methods of allocation	Pre-existing long-term ungrazed plots
	Intervention description	Paired ungrazed and grazed plots (10m x 10m). Grazed plots are grazed at rates typical of area, but not quantified, or whether stock penned in or left open to grazing from wider area.
	Control/comparison description	Ungrazed plots
	Sample sizes	Good sample size – three replicates from each vegetation type. Between 8 and 40 quadrats from large to small scale respectively
	Baseline comparisons	Reported as similar when plots were set up, but no data presented.
	Study sufficiently powered	No power analysis. Good sample size – three replicates from each vegetation type. Between 8 and 40 quadrats from large to small scale respectively. Seven spatial scales (grain size) from 5cm x 5cm to $1m^2$.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	Ground cover in quadrats of different scales, and leaf traits: specific leaf area; leaf dry matter content; leaf nitrogen and carbon content.
	Secondary outcome measures	Vegetation data used to derive indices of spatial variability.

significance)	Follow-up periods	Treatments in place for 20 years (any variation in grazing regimes over this period not reported, but assumed to be fairly consistent). Enough time for grazing-related differences to be detected.
	Methods of analysis	Quadrat species data used to calculate Euclidean distance, averaged among all pair-wise comparisons for each grain size. Aggregated leaf-trait values calculated for each quadrat, weighted by relative abundance of species, and spatial variability calculated at each grain size. These variables tested against grain size for grazed and ungrazed in each grassland. Effects of plant community and grazing treatment tested using mixed effects ANOVA.
Results		Species richness was significantly different between the three plant communities at all grain sizes, with mesic grasslands being most species rich. Evenness and variability of species composition was significantly influenced by community type only at the smallest grain size. Species richness was affected by grazing at grain sizes of $0.25m^2$ and above, but effect differed between communities. In heath and mesic grassland species-richness was higher in the grazed plots, and in xeric calcareous grassland in the ungrazed plots. For the former two communities evenness was lower in grazed plots at scales up to $0.15m^2$, with no difference in xeric grassland. This suggests that grazing probably influences species-richness at the community scale through changes in distribution of relative dominance at the very fine scale. Dissimilarities increased with grazing exclusion in all communities at all but the finest grain size. Variability in species composition decreased more sharply with increasing grain size in grazed plots than in ungrazed. Spatial variability of leaf traits was different between communities, especially at the fine grain sizes, and increased with grazing exclusion, but only for three traits at certain grain sizes. Spatial variability of leaf carbon content was increased by grazing exclusion at most scales, and leaf dry matter content at larger scales. Spatial variability of leaf

		nitrogen increased with grazing exclusion only at fine spatial scales in xeric grassland. The decrease in leaf trait variability with increasing grain size tended to be faster in grazed plots than ungrazed. The findings support the generally held view that in more productive environments moderate grazing pressure can increase plant diversity through reducing competition and increasing heterogeneity of resource availability and establishment opportunities. In low productivity environments other environmental stress limits competition whilst grazing and trampling can have adverse impacts. Spatial variability was however positively affected by long-term grazing exclusion, with grazing having an effect only at small grain-sizes, possibly due to reducing dominance and aggregation. Leaf trait results show grazing exclusion resulting in a coarse grain of leaf trait heterogeneity, which may reflect spatial aggregation of species with similar trait values. The weaker sensitivity of trait values may indicate a degree of functional redundancy (different species with similar trait values). Diversity and spatial heterogeneity of vegetation to grazing exclusion can be partly disconnected, depending on community. Spatial variability could be useful for detecting within community responses to grazing as it is detectable at a very fine scale of sampling and responded similarly across communities.
Notes	Limitations identified by author	
	Limitations identified by review team	Grazing levels not specified, species abundance based on subjective measures of plant cover in quadrats – estimates may be affected by scale of observation.
	Evidence gaps and/pr recommendations for further research	Further research to test the combined responses of diversity and heterogeneity on a larger range of productivity.
	Sources of funding	Not clear. Site part of ILTER- Europe Network

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Quetion	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Deléglise, C., Loucougaray, G & Alard, D (b)
	Year	2011b
	Aim of study	To investigate whether grazing exclusion leads to an increase in patch size and/ or contrast between patches in the studied (subalpine) communities, and relationships between spatial patterns of species and spatial patterns of plant trait values, overall abundance level of community productivity.
	Study design	2
	Quality score	+
	External validity	++
Population and setting	Source population	Subalpine areas of France. Not described in detail
	Eligible population	Likely to be fairly representative of at least part of the range of sub-alpine communities – calcareous grassland, mesic grassland and heath grassland. Described in some detail in terms of characteristic and dominant species.

	Inclusion and exclusion criteria	Experiments based on long-term plots, with original selection not reported. Pairs chosen to be similar in pre-exclosure grazing history, environmental conditions and species composition and diversity.
	Setting	Grassland habitats in Hauts-Plateaux du Vercors Nature Reserve, and heath at Alp d'Huez ski resort, Western Alps, France. Altitude 1600-1800 m asl. High precipitation and snow cover duration.
Methods of allocation to intervention/control	Methods of allocation	Pre-existing long-term ungrazed plots
	Intervention description	Paired ungrazed and grazed plots (10m x 10m). Grazed plots are grazed at rates typical of area, but not quantified, or whether stock penned in or left open to grazing from wider area.
	Control/comparison description	Ungrazed plots
	Sample sizes	Good sample size – three replicates from each vegetation type. Veg height measurements for 199 small quadrats per plot and plant traits from 10 or 20 replicates per plot.
	Baseline comparisons	Ungrazed plots
	Study sufficiently powered	States that the sampling design allowed enough pairs of observations (for calculating correlation between variables at different distance classes) to ensure the power of the test for all distance classes up to 500cm.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	Plant abundance and height. Plant traits are species vegetative height, specific leaf area, leaf dry matter and nitrogen content.
	Secondary outcome	Measures of spatial pattern derived using degree of correlation between values of a

significance)	measures	variable for different distance classes.
	Follow-up periods	Treatments in place for 20 years (any variation in grazing regimes over this period not reported, but assumed to be fairly consistent). Enough time for grazing-related differences to be detected.
	Methods of analysis	Cummulative plot scale abundance of all species belonging to five life forms calculated. Aggregated leaf-trait values calculated for each quadrat, weighted by relative abundance of species. Measure of autocorrelation (Moran's I) calculated to test the degree of correlation between the values of a variable as a function of spatial location. Calculated for different distance classes up to 500cm (see 4.2) Derived measures of Grain and contrast between patches. Effects of treatment, plant community and interactions tested using mixed-effects ANOVA.
Results		Patch size of canopy height ranged from 30cm in heath grasslands to over 1m in xeric grasslands. Grazing had a significant effect in xeric grasslands with a difference in means of 47±23.3 cm vs 123±8.8cm. No significant effects detected for Moran's *I* the intensity of spatial dependence. For species, there was a significant increase in average patch size with grazing exclusion also in xeric grasslands and again no significant effect for Moran's *I.* Grazing exclusion decreased graminoid abundance across all communities, and increased legumes in xeric grasslands. Patchiness of grasses did not differ between grazed and ungrazed plots. Patch size of rosette and non-rosette forbs showed a significant response to grazing exclusion, mainly due to increased patch sized in xeric grasslands, but this did not translate into change in spatial pattern. No significant effects on spatial patterns on plant trait values were detected due to considerable between plot variability within treatment and plant communities. In this study it was found that grazing exclusion resulted in changes in spatial heterogeneity only in the low-productivity xeric community, as indicated by the coarser grain of patchiness. It is suggested that grazing altered spatial spread, rather than intra-

		specific aggregation (intensity of spatial dependence). This may be due to clonal growth, or to weak intra-specific competition under harsh environmental conditions. Study stresses it is important to consider changes in spatial patterns in addition to changes in mean values of vegetation features when assessing the impact of grazing management, as both types of change may occur independently of each other. Long-term grazing exclusion only affected patch size rather than other measures of spatial dependence, and effects significant only in the xeric (low productivity) community. Changes in spatial patterns of species did not support changes in spatial patterns of trait values. Changes in abundance and patch size of life forms were affected by grazing exclusion but this did not correspond to changes in spatial pattern at the scale investigated.
Notes	Limitations identified by author	Scale of the study may affect the likelihood of detecting spatial effects – e.g. dwarf shrub occurred in almost all quadrats in the ungrazed plots that did not allow differences in spatial scale to be distinguished between grazed and ungrazed plots.
	Limitations identified by review team	Details of grazing treatment and variation in it (over time and between sites) not given. Possible confounding of livestock type – cattle grazing with sheep at heath site.
	Evidence gaps and/pr recommendations for further research	To identify ecosystem consequences of observed changes in spatial pattern.
	Sources of funding	Not clear. Site part of ILTER- Europe Network

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. Effect of grazing on biodiversity and other ecosystem services

Study details	Authors	Dennis et al.
	Year	2008
	Aim of study	To investigate the effects of changes in stocking density and differences in species of livestock on foliar arthropods
	Study design	Quantitative experimental 1
	Quality score	+
	External validity	+
Population and setting	Source population	M23, M25, U4, U5 and U20 Altitude 220-500m AOD
	Eligible population	24 plots – mosaic of vegetation types, as described above
	Inclusion and exclusion criteria	Semi-natural acid grassland/mire habitats
	Setting	Glen Finglas, Scotland

Methods of allocation to intervention/control	Methods of allocation	3 paired replicate blocks composed of 8 experimental blocks (2 replicate blocks of 4 plots) Grazing treatments assigned randomly
	Intervention description	Varying stocking rates comprising - commercial density sheep grazing, low density sheep grazing and low density mixed grazing. Sufficient detail to replicate
	Control/comparison description	Control treatment – low sheep grazing (continuation of previous management)
	Sample sizes	Replicated study, sample size of 25 samples points with 5 sub-samples. Analyses based on 247 480 specimens collected over 4 years
	Baseline comparisons	Baseline samples collected Jun-Jul 2002 showed similar mean abundance in areas demarcated for the 4 grazing treatments
	Study sufficiently powered	Not reported, but random, replicated study, therefore likely to give reasonable levels of power
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Arthropod type, abundance and mass
	Secondary outcome measures	N/A
	Follow-up periods	2003-2005

	Methods of analysis	Univariate analyses using REML for abundance of key arthropod groups Abundance data for crane flies analysed using one-way ANOVA
Results		No significant difference between plots recorded on abundance within 6 months, but a significant difference was recorded at 18 and 30 months for all taxa except brachycerans. In 2005, true bugs and beetle numbers were significantly higher in the ungrazed plot cf all grazing treatments In 2004, mean abundance of spiders was significantly higher in the ungrazed and low stocking density (mixed stocking) treatments compared with the commercial stocking density treatment. Cranefly and moth caterpillar numbers significantly related to sheep density interacting with year, but not sheep density alone Arthropod mass was approximately double in the ungrazed and mixed low density treatments cf commercial sheep grazed
Notes	Limitations identified by author	Longer term experiment needed to confirm the true effect of grazing species rather than gross stocking densities on a broader group of arthropods; to confirm whether there is an optimal period for cattle grazing beyond which habitat quality for arthropods diminishes and pulse grazing may be more effective
	Limitations identified by review team	Experimental design robust in many areas, but lack of environmental factors considered – notably temperature and rainfall is a weakness, and reliability of outcome measures was not clear from the report
	Evidence gaps and/pr recommendations for further research	See author limitations above. Also, detailed response of specific arthropod taxa, and analysis of vegetation and arthropod characteristics associated with breeding bird success

	Sources of funding	SEERAD

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____UPLAND_____

Name of Review Sub-topic (if any): _____GRAZING_____

Review Question	Effect of grazing on biodiversity and other ecosystem services
Study Citation	Dennis et al (2008)
Study Design Category	Quantitative experimental 1
Assessed by & when	SUSANNA PHILLIPS 14/11/2012

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 ☐NA	Comments: Glen Finglas, Scotland Unintensified, acid grassland and mire M23, M25, U4, U5 and U20 Vegetation condition not described Altitude 220-500m AOD
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 ☐NA	Comments: 24 plots – mosaic of vegetation types, as described above
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 ☐NA	Comments: Paired blocks arranged adjacent to each other Arthropod sampling - 25 samples points with 5 sub-samples for suction sampling at intersections of grid composed of squares of 40m sides and 20x0.5m transect for net sampling. Samples not selected randomly.

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 ☐NA	Comments: 3 paired replicate blocks composed of 8 experimental blocks (2 replicate blocks of 4 plots) Grazing treatments assigned randomly
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Varying stocking rates comprising - commercial density sheep grazing, low density sheep grazing, low density mixed grazing and no grazing (control). Sufficient detail to replicate
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)	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sheep grazing started January 2003 Cattle only grazed from late summer 2003 Exposure was adequate – as experimental design, except, sheep removed during severe weather (assumed to apply to all plots)
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 ☐NA	Comments: No contamination was 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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sheep removed in severe weather and for dipping – applied to all grazed blocks

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Representative of acid grassland habitats
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Commercial stocking rates, low stocking rates may reflect agri-environment scheme practices, livestock removal may be typical on some sites

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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sample points located using GPS – accuracy not reported D-vac suction sampler and sweep net, and arthropods identified (consistency not reported) Sweep net samples not taken on wet/windy days Samples collected from randomly chosen sample points per visit to avoid bias of different blocks sampled at different times Additional 600 sweep net samples in Nov 2005 to sample crane flies emerging from soil
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)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: All measures reported on, but only 5 suction samples collected in 2005, compared with 25 in other years, samples multiplied by 5 to compensate, may have increased variability
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ ☐+ ☐- ☐NR	Comments: Identification/counting centred on groups sampled effectively by the methods & prominent in diet of upland bird species – no further details provided The study did not account for effect on arthropod populations of other environmental factors – in

	☐NA	particular, annual variation in weather
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)?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Arthropods collected in sampling do not represent total population – addressed to an extent by use of two sampling techniques Mean biomass wet weight per specimen in different taxonomic group taken from 10 randomly chosen suction samples but dependent on ability of sampling to collect individual species
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Sheep grazing treatment started in January 2003, cattle grazing started in late summer 2003, therefore, in particular in earlier surveys, cattle plots were not comparable
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Baseline samples Jun-Jul 2002, repeat samples Apr-Jun 2003, May-Jul 2004 and My-Jun 2005. Longer timescales may have given clearer trends

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 ☐NA	Comments: Some differences expected in vegetation composition, but randomly assigned, replicated study.
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	☐++ ☐+ ☐- ☐NR	Comments: Power calculation not present Replicated study, sample size of 25 samples points with 5 sub-samples. Analyses based on 247 480 specimens collected over 4 years

the expected effect size? Is the sample size adequate?	☐NA	
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Estimates of effect size not reported Mean and sd for species type, plot and year shown graphically
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 ☐NA	Comments: Univariate analyses using REML for abundance of key arthropod groups REML calculated to allow for time lags in treatment implementation Abundance data for crane flies analysed using one-way ANOVA
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 ☐NA	Comments: p-values tabulated
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: Generally robust monitoring and analytical methodologies, but – Arthropods collected in sampling do not represent total population The omission of environmental variables may be significant
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: Detailed description of research, reasonable levels of external validity, taking account of comments in 5.1

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	P Dennis, M R Young, C L Howard & A J Gordon
	Year	1997
	Aim of study	To investigate the relationship between livestock species grazed at different intensities and the arthropod fauna of upland grassland habitats.
	Study design	1 2 -non randomised (DM)
	Quality score	=QA5.1 Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in each plot.
	External validity	=QA5.2 Methods given in great and repeatable detail
Population and setting	Source population	Species-poor Nardus stricta grassland on Blackdean Curr at 450-500m in the Cheviot Hills

	Eligible population	Species-poor Nardus stricta grassland
	Inclusion and exclusion criteria	
	Setting	
Methods of allocation to intervention/control	Methods of allocation	Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in each plot.
	Intervention description	
	Control/comparison description	A further treatment had been lightly grazed for 6 weeks during 1991 and 1992 but all grazing had ceased by autumn 1992. This was used as a short-term ungrazed control, ave sward height 8-12cm.
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each	Primary outcome measures	

outcome and significance)	Secondary outcome measures	
	Follow-up periods	
	Methods of analysis	
Results		32 species of carabid and staphylinid beetles were captured in high enough numbers to investigate. 8 species did not respond to the variation, 10 spp correlated with heavier grazing intensity, 6 spp correlated with taller mean veg height, 8 spp correlated with low grazing or ungrazed areas. 'A rotation of varied management over time, different combinations of grazers and varied grazing intensities would encourage a wider diversity of Coleoptera by creating a mosaic of structurally varied grassland patches'.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NR

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	The response of epigeal beetles to varied grazing regimes on upland Nardus stricta grasslands. P Dennis, M R Young, C L Howard & A J Gordon Journal of Applied Ecology (1997) 34, pp433-443
Study Design Category	1 2 non-randomised (DM)
Assessed by & when	Alison Hiles 6/3/2013

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 ☐NA	Comments: Species-poor Nardus stricta grassland on Blackdean Curr at 450-500m in the Cheviot Hills
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 ☐NA	Comments:
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 ☐NA	Comments:

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 ✔ ☐NA	Comments:
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 2 livestock treatments of sheep and sheep+cattle so that each maintained 2 average, between-tussock heights of 4-5cm and 6-7cm. A further treatment had been lightly grazed for 6 weeks during 1991 and 1992 but all grazing had ceased by autumn 1992. This was used as a short-term ungrazed control, ave sward height 8-12cm. All treatments replicated twice to give a total of 10 plots, varying in size to accommodate 6 yearling steers in the mixed livestock treatments from June to August each year. Variation in size allows for numbers of sheep to be similar across livestock treatments. Sheep numbers per plot adjusted weekly to maintain target sward heights.
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)	✔ ☐++ ☐+ ☐- ☐NR ☐NA	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 ☐NA	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional	☐++ ☐+ ☐-	Comments:

interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR ✔ ☐NA	
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK resource.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Nardus dominated fells are typical of that type across the English uplands
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

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?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in 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)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	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)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

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 ☐NA	Comments: A further treatment had been lightly grazed for 6 weeks during 1991 and 1992 but all grazing had ceased by autumn 1992. This was used as a short-term ungrazed control, ave sward height 8-12cm.

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?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.3 Were the estimates of effect size given or calculable?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	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?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
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 ☐NA	Comments:
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: Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was

		calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in each plot.
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: Methods given in great and repeatable detail

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	Peter Dennis, Mark R Young, Christopher Bentley
	Year	2001
	Aim of study	The effects of varied grazing management on epigeal spiders, harvestmen and pseudoscorpions of Nardus stricta grassland in upland Scotland
	Study design	2
	Quality score	=QA5.1 Continuous sampling with pitfall traps; monthly suction samples; visual search for spiders' webs. Exact details of all methods given and results produced of 21,758 individuals of 83 spp. 'Even given limited changes to botanical species composition in Nardus-dominated grasslands, prey abundance could not be entirely excluded as a factor.'
	External validity	=QA5.2x No livestock, sheep only or sheep with cattle, grazed to maintain average, between-tussock heights of 4.5cm or 6.5cm. These sward heights were maintained by continuous but varied stocking rates of sheep between May and October from 1991- 1994. All treatments replicated twice to give a total of 10 plots, enclosed with post and wire fencing across 22ha of Nardus stricta-dominated grassland lying on the summit and adjacent ridges of Blackdean Curr. Pitfall traps; monthly suction samples; visual search for spiders' webs used to collect species. Exact details of all methods given. Data compared with stocking rate, botanical spp composition and vegetation structure.

Population and setting	Source population	
	Eligible population	
	Inclusion and exclusion criteria	
	Setting	
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	
	Secondary outcome measures	
	Follow-up periods	
	Methods of analysis	

Results		'There was a significant effect of grazing treatment on the total number of arachnid spp and linyphiid spp and total individuals, all showing the same trend of most spp in ungrazed, taller, sheep grazed swards. Fewest spp were associated with both the sheep + cattle grazed treatments' 'Significantly more webs were counted in ungrazed than other treatments in both 1993 and 1994'
Notes	Limitations identified by author	'Even given limited changes to botanical species composition in Nardus-dominated grasslands, prey abundance could not be entirely excluded as a factor because this too can respond to the architecture of individual plant species.'
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NR

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	The effects of varied grazing management on epigeal spiders, harvestmen and pseudoscorpions of Nardus stricta grassland in upland Scotland Peter Dennis, Mark R Young, Christopher Bentley Agriculture, Ecosystems and Environment 86 (2001) pp39-57
Study Design Category	2
Assessed by & when	Alison Hiles 6/3/2013

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 ☐NA	Comments: Species-poor Nardus stricta grassland on Blackdean Curr at 450-500m in the Cheviot Hills
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 ☐NA	Comments:
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 ☐NA	Comments:

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 ☐NA	Comments:
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: No livestock, sheep only or sheep with cattle, grazed to maintain average, between-tussock heights of 4.5cm or 6.5cm. These sward heights were maintained by continuous but varied stocking rates of sheep between May and October from 1991- 1994. All treatments replicated twice to give a total of 10 plots, enclosed with post and wire fencing across 22ha of Nardus stricta-dominated grassland lying on the summit and adjacent ridges of Blackdean Curr. Pitfall traps; monthly suction samples; visual search for spiders' webs used to collect species. Exact details of all methods given. Data compared with stocking rate, botanical spp composition and vegetation structure.
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)	☐++ ☐+ ☐- ☐NR ✔ ☐NA	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 ✔ ☐NA	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups?	☐++ ☐+ ☐-	Comments:

Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR ✔ ☐NA	
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Scottish border with England. Nardus-dominated moorland typical of upland England as well
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

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?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	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)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Continuous sampling with pitfall traps; monthly suction samples; visual search for spiders' webs. Exact details of all methods given and results produced of 21,758 individuals of 83 spp
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the	✔ ☐++ ☐+ ☐-	Comments:

variables/measurements used?	☐NR ☐NA	
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)?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

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 ☐NA	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.	☐++ ☐+ ☐-	Comments:

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	✔ ☐NR ☐NA	
4.3 Were the estimates of effect size given or calculable?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	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?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
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 ☐NA	Comments:
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: Continuous sampling with pitfall traps; monthly suction samples; visual search for spiders' webs. Exact details of all methods given and results produced of 21,758 individuals of 83 spp. 'Even given limited changes to botanical species composition in Nardus-dominated grasslands, prey abundance could not be entirely excluded as a factor.'
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: No livestock, sheep only or sheep with cattle, grazed to maintain average, between-tussock heights of 4.5cm or 6.5cm. These sward heights were maintained by continuous but varied stocking rates of sheep between May and October from 1991-1994. All treatments replicated twice to give a total of 10 plots, enclosed with post and wire fencing

nationally (i.e. habitat, species)?		across 22ha of Nardus stricta-dominated grassland lying on the summit and adjacent ridges of Blackdean Curr. Pitfall traps; monthly suction samples; visual search for spiders' webs used to collect species. Exact details of all methods given. Data compared with stocking rate, botanical spp composition and vegetation structure.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?

Study details	Authors	P Dennis, R J Aspinall, Iain J Gordon
	Year	2001
	Aim of study	Spatial distribution of upland beetles in relation to landform, vegetation and grazing management.
	Study design	2
	Quality score	=QA5.1 Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in each plot.

		As Dennis et al (1997)
	External validity	=QA5.2 Methods given in great and repeatable detail
Population and setting	Source population	
	Eligible population	
	Inclusion and exclusion criteria	
	Setting	
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	
	Secondary outcome measures	

	Follow-up periods	
	Methods of analysis	
Results		'The analyses identified Carabus problematicus and Olophrum piceum as the species most sensitive to grazing management.....Both these species related to lower grazing intensity, expressed either as taller grass or lower stocking rates. However, C. problematicus occurred in Nardus grazed by sheep and not by cattleand sheep on the higher slopes, whereas O. piceum occurred in taller vegetation that had been ungrazed for 2 years.' 'To summarize, the large-scale spatial associations related mainly to landform pattern whereas the smaller ones corresponded to the modification of the vegetation through grazing management
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Scottish Executive Rural Affairs Department Flexible Fund

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Upland _____

Name of Review Sub-topic (if any): _____Moorland grazing_____

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Spatial distribution of upland beetles in relation to landform, vegetation and grazing management. P Dennis, R J Aspinall, Iain J Gordon Basic Applied Ecology 3 pp183-193 (2002)
Study Design Category	2
Assessed by & when	Alison Hiles 8/3/2013

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 ☐NA	Comments: Species-poor Nardus stricta grassland on Blackdean Curr at 450-500m in the Cheviot Hills. As Dennis et al (1997)
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 ☐NA	Comments: As Dennis et al (1997)
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 ☐NA	Comments:

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 ✔ ☐NA	Comments:
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: As Dennis et al (1997)
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)	✔ ☐++ ☐+ ☐- ☐NR ☐NA	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 ☐NA	Comments:
2.5 Were any 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 ✔ ☐NA	Comments:

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK resource.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Nardus dominated fells are typical of that type across the English uplands
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

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?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in 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)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative	✔ ☐++ ☐+	Comments:

effects assessed by the variables/measurements used?	☐- ☐NR ☐NA	
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)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: To apply the distance statistics method, data were used on the trap abundance of selected species of ground and rove beetles sampled in individual pitfall traps, 120 in total, that were spatially referenced by satellite telemetry and corroborated with map and compass/protracotr
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

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 ☐NA	Comments: 6 species selected for further analyses
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.	☐++ ☐+ ☐-	Comments:

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	✔ ☐NR ☐NA	
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	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?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
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 ☐NA	Comments:
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: Effects of the treatments were characterised by measures of the botanical composition, general vegetation height and total grazing pressure per treatment. Botanical data obtained from quadrats in June and September of 1993&4 at 24 locations in each plot to indicate the total floristic diversity. The height profile of all vegetation, including tussocks within a 6m radius of the 12 pitfall traps was measured at 40 points. The mean veg height of each of the 10 plots was calculated from the 480 measurements thus recorded in each plot. Beetle assemblage was sampled with 12 pitfall traps located randomly along existing transects marked across each of the 10 plots. Variable efficiency of capture of different

		species by pitfall so only dealt with the most frequently trapped species and combined the catches from all the traps in each plot.
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: Methods given in great and repeatable detail

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	(a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services

Study details	Authors	David JT Douglas, Darren M Evans and Stephen M Redpath
	Year	2008
	Aim of study	To test the hypothesis that on intensively grazed moorland, breeding Meadow Pipits forage for nestling food where arthropod prey are most readily available, and therefore that foraging site choice is a function of prey abundance and vegetation structure
	Study design	Quantitative observational correlation study 2
	Quality score	-
	External validity	+
Population and setting	Source population	Soligenous mire/grassland mosaic (NVC types: M25, M23, U4)
	Eligible population	Plots within above mosaic used but selection of plots not described
	Inclusion and exclusion criteria	

	Setting	Scotland (Glen Finglas)
Methods of allocation to intervention/control	Methods of allocation	Samples were taken at points where the pipits took off after feeding and also at control points where no feeding had been observed.
	Intervention description	
	Control/comparison description	Control points were those at which no feeding was observed
	Sample sizes	19 nests were watched. 4 'foraging sites' were identified per nest. Each 'foraging site' was paired with a control site where birds had not been observed to forage.
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Sward height, sward density and arthropod biomass were recorded at each sample point. Arthropods were also identified as far as Order and the biomass of each Order was recorded.
	Secondary outcome measures	
	Follow-up periods	One sampling period, one year after different treatments introduced into study plots.
	Methods of analysis	
Results		The paper concludes that foraging sites had lower vegetation height and higher total arthropod biomass than control sites and that therefore, in heavily grazed upland systems, Meadow Pipits select foraging sites that optimise total food abundance and accessibility. However, the difference in the cumulative biomass of selected prey types

		between foraging sites and control sites was not statistically significant.
Notes	Limitations identified by author	Biases associated with recording prey items (smaller items may be under-recorded or missed)
	Limitations identified by review team	No evidence is provided that the sample points studied (location of point from which meadow pipit took off) were related to where foraging took place. Take-off point might have been more closely related to good take-off conditions (freedom from obstruction or good visibility) The different grazing treatments had only been in place on the sample plots for one year. Vegetation and invertebrate communities are likely to change further over time.
	Evidence gaps and/or recommendations for further research	To be able to draw firm conclusions about the effects of different grazing regimes on meadow pipit populations it would be necessary to: make a direct link between sample points and foraging activity carry out longer term studies of the impacts of different grazing regimes upon invertebrate communities (especially abundance of prey species) carry out longer term studies of the results of different grazing regimes upon vegetation structure investigate the effects of sward structure and predation risk on the pipits determine whether there is an optimal grazing regime that delivers both maximum prey abundance and optimal vegetation structure – if not, then the relative importance of the two variables needs to be further investigated
	Sources of funding	Scottish Executive Environment and Rural Affairs Department

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

Name of Evidence Review: Upland

Name of Review Sub-topic (if any): Grazing

Review Question	(a)
Study Citation	David J T Douglas, Darren M Evans, Stephen M Redpath (2008): Selection of foraging habitat and nestling diet by Meadow Pipits *Anthus pratensis* breeding on intensively grazed moorland
Study Design Category	Quantitative Observational Correlation Study 2
Assessed by & when	Jean Johnston, 7/11/12

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: A list of the 3 NVC types present in the general area is given (M25, M23, U4).
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: The vegetation present in the study plots is not described any further.
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: The method for selection of the study plots is not described. This had been done for other purposes before this study began.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++	Comments: Selection of study sites within plots (sites used for foraging vs sites not used for foraging) was made by following the pipits. This should be free from observer bias.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Vegetation height and density and arthropod biomass are sound variables to look at. Vegetation composition was not considered.
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: Not relevant –the birds' choice of foraging sites would not affect the variables that were considered.
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:
2.5 Is the setting applicable to the UK?	☐++	Comments: This is a UK (Scottish) study

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: Measures were objective. Vegetation height and density are related.
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:

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: It is noted in the paper that pipits generally landed in one place, walked along the ground for some distance and then returned to the nest. It is also noted that parent birds were multiple prey loaders (carrying 1-6 items). Foraging sites were defined as the last observed location of the pipits before they flew back to the nest. No evidence is given as to whether the food was usually collected from this point. It might be, for example, that the last point is chosen as a good take-off area (e.g free from obstruction and/or good visibility)
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments:
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: This study was completed when the different plot treatments had only been in place for one year.

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:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.3 Were the analytical methods appropriate? Were important differences in follow-up time	☐++ ☐+ ☐-	Comments:

and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐NR ☐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 ☐NA	Comments:
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: The lack of evidence over whether the last observed location of the pipit represents the foraging area is a significant flaw.
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: If not flawed, could probably be generalised to short-term responses in small plots of rough grassland/soligenous mire mosaics in the British Uplands. However, this does not necessarily apply to other habitats such as ombrogenous mires, heaths, montane habitats or rocky habitats – and does not consider responses on a landscape scale. As a short term study, it necessarily takes no account of longer term habitat changes eg from grassland to heath.