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 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	Evans, R. (1977) Overgrazing and soil erosion on hill pastures with particular reference to the Peak District. Journal of the British Grassland Society, 32, 65-76. Evans, R., 2005. Curtailing grazing-induced erosion in a small catchment and its environs, the Peak District, Central England. Applied Geography 25, 81-95
Study Design Category	2, 3
Assessed by & when	D Martin 6/11/12, April 2013

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 upland catchments. Only study area is described but there is some reference to sheep-initiated erosion in other parts of upland Britain.
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 study drainage basin is said to be typical of parts of the Peak District. The habitats described are typical of upland hill grazing over podzols and peaty podzols and gleys.
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 whole sub-catchment is described in detail in terms of topography and vegetation. Slope erosion processes and areas of bare soil were observed to be present.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Census study – the prevailing agricultural grazing conditions are measured, and the occurrence of bare ground arising from other processes. The study area was chosen from aerial photographs as typical of the area, but selection was subjective.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: The basis for the assumed relationship between vegetation, slope and bare ground. The relationship between sheep and erosion scars had previously been described (Thomas 1965)
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: Observational study at one site – but divided into areas based on morphology and vegetation. Different causes of bare ground identified and recorded. Study continued for a number of years with regular monitoring visits to assess recovery, following a route linking the identified scars.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes – study site is in UK and at typical hill grazing altitude and vegetation types.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Outcome measures include estimates of bare soil, estimated on transects and sample areas. Means of values from two observers. Erosion rates measured at thirty-one sites – subjectively chosen to include a range of sizes, weighted towards larger erosion scars. Erosion rates measured objectively using pins and nail markers, and downslope collection of soil. Ongoing monitoring visits at eight times between 1974 and 2001, with scars photographed from a fixed point at each visit. Area of bare ground was measured from photos using a scaled grid.
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?		Comments: Yes, erosion rates and expansion of scars

Were all important positive and negative effects assessed?	☐+	assessed. Vegetation colonisation measured during a later follow-up period.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Yes – direct measures of erosion
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: Measured over a two-year period. Only one site. The erosion scars were originated before the study, presumably at different times, and possibly several years before. Monitoring continued by observation for a further 33 years, with eight visits between 1974 and 2001.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Initial study over two years so may not reflect range of erosion rates during different periods of weather e.g. fewer days of ground frost. Longer term effects were assessed in a follow-up 6 years later in 1974, but the treatment (i.e. grazing levels) and climatic effects will not have stayed constant over this time. Subsequent paper reports regular visits to 2001.

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. Sample size is reasonably large (31 erosion areas) and 108 sheep counts over the two year period.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: relationship between sheep numbers and bare ground is the only one explored analytically. Climatic trends considered in the re-colonisation study, but not statistically.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: yes, but limited. Regression equations for sheep density and bare ground. Variance explained is low, but improved by removing bare ground obviously due to human activity, such as burns and tracks. Area of bare ground plotted over time in the continuation paper.
4.4 Was the precision of the intervention effects given or calculable? Is association	☐++	Comments: P values given for regression expressions.

meaningful? 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: Census approach on one sub-catchment. Detailed observations of sheep occupancy, and objective measures of erosion.
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: Applicable to immediate upland area – Peak District, and likely to have some read-across to upland areas of similar soils and topography in other UK upland areas.

Thomas, T.M. (1965) Sheet erosion induced by sheep in the Pumlumon (Plymlimon) area, Mid-Wales. In: *Rates of Erosion and weathering in the British Isles*. Inst British Geographers, Geomorphological Symposium pp11 -14

Evidence Table

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

Study details	Authors	Darren M Evans, Stephen M Redpath, Sharon A Evans, David A Elston and Peter Dennis
	Year	2005
	Aim of study	To investigate the effect of livestock grazing density on meadow pipit egg size
	Study design	Randomised controlled trial
	Quality score	+
	External validity	+
Population and setting	Source population	Soligenous mire/grassland mosaic (NVC types: M25, M23, U4)
	Eligible population	3.3ha 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	Randomised
	Intervention description	Plots grazed by: 2.72 ewes/ha, 0.91 ewes/ha, 0.61 ewes/ha and ungrazed

	Control/comparison description	
	Sample sizes	6 replicates of each treatment
	Baseline comparisons	Baseline recorded
	Study sufficiently powered	?
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Egg size
	Secondary outcome measures	
	Follow-up periods	Treatments only in place for less than 1 year (autumn 2002 to spring 2003)
	Methods of analysis	
Results		On this soligenous mire/grassland site, after less than 1 year, intensively grazed plots (2.72 ewes/ha) contained nests with the smallest eggs and extensively grazed plots (0.61 ewes/ha) contained nests with the largest eggs. Ungrazed plots contained eggs with smaller eggs than lightly grazed plots. The study found, from 82 nests measured, that after less than 1 year, intensively grazed plots (I) contained nests with the smallest eggs and extensively grazed plots (III, grazed at 2 ewes only at time of study) contained nests with the largest eggs. Ungrazed plots contained eggs with smaller eggs than lightly grazed plots. There was no significant effect of breeding density, laying date or clutch size on egg volume. No effect of egg size or grazing treatment on fledging success was found, which may be due

		to compensatory mechanisms or effects becoming apparent post-fledging. The mechanisms of effect of grazing and egg size remain unclear. Whilst it is likely that food availability is important, grazing may also affect territory size and hence parental quality, and nest microclimate.
Notes	Limitations identified by author	
	Limitations identified by review team	Treatments only in place for very short time, so longer term effects, including those arising from changes to vegetation composition not considered. This is only a plot study so landscape scale effects not considered. This study only considers sheep grazing and not the effects of other domestic or wild herbivores.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Scottish Executive Environment and Rural Affairs Department

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: Upland

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

Review Question	(a)
Study Citation	Livestock grazing affects the egg size of an insectivorous passerine
Study Design Category	Randomised controlled trial
Assessed by & when	Jean Johnston, 9/1/13

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 vegetation is not described in this paper but I assume it is the same as in Evans.JAE.2006 where a list of the 3 NVC types present in the area is given (M25, M23 and U4) but no further description is provided and the proportions of each type are not given.
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 plots are not individually 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?	☐+	Comments: Measurements were taken at all nests with eggs that were found in the plots. Nests with eggs that had already hatched were excluded – no data is given as to whether these early-hatched nests were evenly dsitributed between plots. Otherwise, selection should be free from bias.

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: Randomised
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Sheep grazing densities are well described, but breeds not given
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments:
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	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: This is a UK (Scottish) study
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	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?	☐++	Comments:
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐-	Comments: Treatment had been in place for less than a full year (autumn 2002 to spring 2003)

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: There was significant variation between plots before the treatments began but the analysis allows for this.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐-	Comments: Power calculations are not provided. Sample sizes are

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?		not especially large (49 nests in 2003, with between 7 and 16 nests per treatment)
4.3 Were the estimates of effect size given or calculable?	☐++	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments:
4.5 Was the precision of the intervention 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: Main concern is short term nature of study.
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: Description of vegetation types is vague, but findings likely to be reasonably applicable to other similar sized plots of similar habitats

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	Darren M Evans, Stephen M Redpath, Sharon A Evans, David A Elston, Charles J Gardner, Peter Dennis and Robin J Pakeman Plus e-mail update Robin Pakeman to David Martin 9/10/12
	Year	2006
	Aim of study	To investigate the effects of different grazing regimes on meadow pipit abundance
	Study design	Randomised controlled trial
	Quality score	- Changed to + as part of robust Glen Finglas expt
	External validity	+
Population and setting	Source population	Soligenous mire/grassland mosaic (NVC types: M25, M23, U4)
	Eligible population	3.3ha 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	Randomised
	Intervention description	Plots grazed as follows: (i) 2.72 ewes/ha, (ii) 0.91 ewes/ha, (iii) 0.61 ewes/ha plus 2 cows and 2 calves for up to 4 weeks (to give overall offtake similar to 0.91 ewes/ha) (iv) Ungrazed
	Control/comparison description	
	Sample sizes	6 replicates of each treatment
	Baseline comparisons	Baseline not recorded
	Study sufficiently powered	?
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Number of breeding territories estimated using CBC methodology
	Secondary outcome measures	
	Follow-up periods	3 years in initial paper, an additional 6 years in e-mail update

	Methods of analysis	
Results		After 2 years of cattle grazing (3 years into the experiment), significantly more pipit breeding territories were found in low intensity mixed livestock grazing plots (treatment iii) than in the other plots. However, this effect did not last. The data presented in the e-mail update shows the number of breeding territories in plot (iii) declining back down to similar levels to the other plots by the 8th year of cattle grazing. There are significant variations in the results from year to year. In the 8th year of cattle grazing, the difference between treatments is reasonably small, varying between just over 2 territories per plot and just over 3 territories per plot. Looking at the longer term results presented in the e-mail update, there is considerable variation in the results from year to year and it is difficult to draw conclusions from this study. In his e-mail Robin Pakeman states "treatment 4 (no grazing) is almost always the worst" but due to the large year-on-year variation and small magnitude of differences between plots in some years, it is difficult to judge the significance of this.
Notes	Limitations identified by author	
	Limitations identified by review team	Baseline not recorded Livestock breeds not described Unsure of significance of observer bias in CBC method for estimating number of breeding territories Detailed vegetation types of individual plots not described. Can't rule out that non-grazing related habitat differences may be affecting results (e.g Pipits might do better in wetter plots in dry years). No data is presented on whether the vegetation composition is changing in the different plots over the course of the experiment.

		This is a plot study and it cannot take into account landscape scale responses
	Evidence gaps and/pr recommendations for further research	Comparisons should be done over large sites under different grazing regimes.
	Sources of funding	Scottish Executive Environment and Rural Affairs Department

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: Upland

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

Review Question	(a) and (g)
Study Citation	Low intensity, mixed livestock grazing improves the breeding abundance of a common insectivorous passerine. Darren M Evans, Stephen M Redpath, Sharon A Evans, David A Elston, Charles J Gardner, Peter Dennis and Robin J Pakeman Updated by e-mail from Robin Pakeman to David Martin 9/10/12
Study Design Category	Randomised controlled trial
Assessed by & when	Jean Johnston 9/1/13

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 vegetation is not described in this paper but I assume it is the same as in Evans.JAE.2006 where a list of the 3 NVC types present in the area is given (M25, M23 and U4) but no further description is provided and the proportions of each type are not given.
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: Plots are not individually 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?	☐++	Comments: All meadow pipit territories were mapped within each plot, so there should be no bias within plots.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Randomised
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Breeds of grazing animals not given
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: Cattle grazing was only 2 cows and 2 calves for up to 4 weeks. This is not representative of normal practice or behaviour on larger sites
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: This is a UK (Scottish) study
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐-	Comments: See 2.3 above

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?	☐-	Comment Estimation of number of breeding territories by Common Birds Census methods may be subject to observer bias
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: Only number of breeding territories is considered.
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: No evidence is given that number of breeding territories is a good surrogate for breeding pipit abundance, though it does seem a reasonable assumption
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: The study described in the paper covers 3 years, the e-mail provides a further 6 years' data. Although this is a total of 9 years, this is still a short time in terms of rates of change of upland habitats.

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: There were difference between treatments at baseline but this is considered in the analysis

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: No power calculation is provided
4.3 Were the estimates of effect size given or calculable?	☐++	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments:
4.5 Was the precision of the intervention 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: Unsure how much observer bias may be affecting the results. Detail of vegetation in plots is not given. DM – 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: Other than the concerns above, results likely to be reasonably applicable to other similarly sized plots of similar vegetation types

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	(g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differently

Study details	Authors	Darren M Evans, Stephen M Redpath, David A Elston, Sharon A Evans, Ruth J Mitchell and Peter Dennis
	Year	2006
	Aim of study	To examine the effects of livestock grazing on field vole abundance
	Study design	Randomised controlled trial
	Quality score	+
	External validity	+
Population and setting	Source population	Soligenous mire/grassland mosaic (NVC types: M25, M23, U4)
	Eligible population	3.3ha plots within above mosaic used but selection of plots not described
	Inclusion and exclusion criteria	
	Setting	Scotland (Glen Finglas)
Methods of allocation	Methods of allocation	Randomised

to intervention/control	Intervention description	4 treatments: conventional stocking rate (3 ewes/ha), one third conventional stocking rate (sheep), one third conventional stocking rate (sheep and cattle), ungrazed
	Control/comparison description	
	Sample sizes	6 replicates of each treatment
	Baseline comparisons	Baseline recorded
	Study sufficiently powered	?
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Presence or absence of droppings in 5 quadrats per plot.
	Secondary outcome measures	
	Follow-up periods	2 years of follow-up
	Methods of analysis	
Results		On this soligenous mire/grassland site, after 2 years, a significantly higher abundance of voles was found in the extensively grazed mixed treatment (cattle and sheep) than in the extensively grazed treatment that contained only sheep. after 2 years, a significantly higher abundance of voles was found in the extensively grazed mixed treatment (III, cattle and sheep) than in the extensively grazed treatment that contained only sheep (II), and particularly the intensively grazed treatment (I). Densities were highest in the ungrazed treatment. The results suggest that low

		intensity and mixed livestock grazing could help mange vole populations in establishing woodland, whilst also improving availability to raptors through increased heterogeneity of vegetation.
Notes	Limitations identified by author	Was only a short-term study. Takes no account of longer term habitat changes or the interactions that treatments may have with vole population cycles.
	Limitations identified by review team	Little detail provided on exact grazing regimes in place. No detail of breeds.
	Evidence gaps and/pr recommendations for further research	Longer term studies needed to examine the interactions of grazing regimes and vole population cycle Landscale-scale responses need to be considered The effects of high vole populations on tree regeneration need to be quantified.
	Sources of funding	Scottish Executive Environment and Rural Affairs Department

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: Upland

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

Review Question	(g) and (h)
Study Citation	Darren M Evans, Stephen M Redpath, David A Elston, Sharon A Evans, Ruth J Mitchell and Peter Dennis (2006): To graze or not to graze? Sheep, voles, forestry and nature conservation in the British uplands. Journal of Applied Ecology, 43, 499-505
Study Design Category	Randomised Controlled Trial 1
Assessed by & when	Jean Johnston, 28/11/12

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: A list of the 3 NVC types present in the area is given (M25, M23 and U4) but no further description is provided and the proportions of each type are not given
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 plots are not individually 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?	☐++	Comments: 5 randomly selected points are used within each plot for recording

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: Randomised
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐-	Comments: Stocking rates are given but no indication of whether these are maxima or averages. Plot sizes are 'approximate' and it is difficult to see how a stocking rate of exactly 1 ewe/ha can be achieved on a 3.3ha plot, unless there is some seasonality in the grazing. No details of actual numbers of animals and dates present (e.g. were they removed for any time e.g. for veterinary treatment, clipping, lambing etc ?) Also no detail provided on breeds or the composition of the 'sheep and cattle' treatment (what proportion of cattle and what number of cows were considered equivalent to a sheep?). However, these rates are described better in other papers in the same series.
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: As above, little detail is provided on actual stocking regime
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: Although little detail is provided, it seems unlikely that there would have been bias.
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	☐NR	Comments: Not reported but seems unlikely

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: This is a UK (Scottish) study.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The stocking rates quoted are within the usual range. As acknowledged by the paper, seasonal variation in stocking is usual and not accounted for in this study.

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: Presence or absence of droppings was used. A study is quoted that showed that vole indices are linearly related to actual vole numbers.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐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)?	☐++	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes
3.6 Was the post-treatment time interval meaningful?	☐-	Comments:

Was the interval long enough to assess long-term effects?		No. It is acknowledged in the paper that this was only a short term study. It is likely that vole numbers would continue to change should the experiment be continued.

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: There was no significant difference between plots pre-treatment
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: No power calculation presented. However, sample size seems reasonable.
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: Main query is over whether plots were indeed comparable vegetation types. If the proportions of the 3 NVC types given were significantly different, this would affect the results. However, there were 6 replicates of each treatment so this is not likley to be a large effect.
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: Can probably be generalised to short-term responses in plots of other 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 landscape-scale responses. As a short term study, it necessarily takes no account of longer term habitat changes eg from grassland to heath.

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	Evans, R
	Year	1997, 2005
	Aim of study	To assess how erosion scars are initiated; what processes were acting on the bare soil; and if bare soil was expanding in the area. To describe the sequence of re-colonisation of bare soil over a period of 32 years.
	Study design	2, 3
	Quality score	+, +
	External validity	+, +
Population and setting	Source population	Source population is upland catchments. Only study area is described but there is some reference to sheep-initiated erosion in other parts of upland Britain.
	Eligible population	The study drainage basin is said to be typical of parts of the Peak District. The habitats described are typical of upland hill grazing over podzols and peaty podzols and gleys.
	Inclusion and exclusion	Slope erosion processes and areas of bare soil present.

	criteria	
	Setting	Hey Clough in the headwaters of the Derwent, North Peak district. Between 253 and 422 m ASL.
Methods of allocation to intervention/control	Methods of allocation	Selected from aerial photograph study, seen as typical of area, but selection subjective.
	Intervention description	The prevailing agricultural grazing conditions are measured, and the occurrence of bare ground arising from other processes.
	Control/comparison description	None
	Sample sizes	One study area. Bare ground measured in thirty 0.5m quadrats. A similar number of scars selected for detailed erosion measurements. Ongoing monitoring visits to 32 erosion scars.
	Baseline comparisons	Vegetation mapped at start and extent of soil erosion sample marked.
	Study sufficiently powered	No power analysis. Sample size is reasonably large (31 erosion areas) and 108 sheep counts over the two year period
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Occurrence and extent of bare ground. Erosion rates and change in bare ground
	Secondary outcome measures	Rates or re-colonisation over a 32 year period.
	Follow-up periods	Measured over 2 years. Re-colonisation study 6 years later (1974). Further monitoring at eight intervals over 27 years.

	Methods of analysis	
Results		Overall the area of bare ground was small (2% of basin) but some large scars were prominent. About 35% of bare soil was classed as sheep scars with other areas accounted for by tracks, burns, gravity scars and discrete patches in old heather. Erosion found to be taking place primarily within Agrostis-fescue swards and also heather moor, but not on peat covered slopes with cotton-grass. Erosion mainly from scars which were expanding by 9.3 mm per year with loss of up to 34 t ha per year of soil. The grass swards supported a higher density of sheep. With mat grass and cotton-grass having the lowest densities. The regression equations suggest that bare ground is initiated at densities of 1 sheep per 0.53 ha (or per 0.58 ha if human influenced bare ground removed) (1.89 and 1.72 sheep per ha). Erosion measurement shows that sheep disturbance of soil cliffs accounted for 22% of erosion movements but 77% of retreat of vegetation. By 1974 most scars had been re-colonised by vegetation despite short growing seasons in preceding years. This was related to a marked decrease (by 25-30%) in sheep numbers in 1968, and not obvious climatic factors. It is suggested that erosion potential thresholds are below the carrying capacity in terms of productivity of the better grassland vegetation. By 2001, thirty-six years after monitoring began, only eight of the original thirty-two scars remained visible. The rate of recolonisation was fastest over the initial ten-year period. At Back Tor expansion continued for longer, although the rate of expansion slowed, with re-colonisation not starting until the peat and organic soil horizons had largely been lost, with an estimated depth of around 450mm total between the peat and leached horizon below. Estimated loss rate was 13mm per year. The remaining scars were still used by sheep for shelter and shade. Although Back Tor was subsequently fenced, the reduction in rate of bare ground expansion and subsequent colonisation coincided with reductions in grazing pressure. It is postulated that bare ground is initiated at summer grazing intensities of between 2.5 and 5 sheep ha^{-1} (0.25-

		0.5 LU ha^{-1}) on short grass and 0.5 sheep ha^{-1} (0.05 LU ha^{-1}) on peat, with a reduction of 30% on these rates allowing recolonisation to start, at least on mineral soils. On eroding peat, recolonisation may not take place until the mineral soil B horizon, which is less acid and more nutrient rich, is exposed. The effects of different stocking rates may on recolonisation may vary with growing season. It is noted that *Calluna* and *Vaccinium* had colonised previously eroded areas, but were absent in *Molinia* dominated areas. Rowan and birch saplings had established in areas of dense heather.
Notes	Limitations identified by author	
	Limitations identified by review team	Single area study, limited analysis. Climatic effects are considered but not analysed as there is no run of data on erosion rates.
	Evidence gaps and/pr recommendations for further research	Study of change in erosion scars in other areas, and identification of critical stocking densities.
	Sources of funding	None given

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

Name of Evidence Review: _____

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

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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?
Study Citation	Ferriera, L. M. M., Oliván, M., Rodrigues, M. A. M., García, U. & Osoro, K. (2005). Estimating diet selection of goats and sheep grazing on gorse-heathland vegetation with areas of improved pasture. In: Silvopastoralism and sustainable land management, Ch 33.
Study Design Category	2
Assessed by & when	D Martin 18/11/12

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: Not in detail – heatlands in LFA of NW Spain.
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 is gorse heath with patches of grass and clover. Representativeness not stated, but assumed to be.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	Comments: Selection not described – likely to be subjective, but chosen to have typical characteristics and vegetation composition. An area of 5 ha sown grassland in the 22 ha plot.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NA	Comments: Both sheep and goats grazed in same area at same numbers (42 of each)
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Yes – availability of patches of different vegetation types which are likely to vary in palatability and likelihood of selection by grazing animal.
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: No replication, may be interaction between the grazing animals – species not tested individually.
2.5 Is the setting applicable to the UK?	☐+	Comments: NW Spain, but some similarities to western heath of SW England.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Objective measures of diet composition from alkane concentrations in faeces. Use of the method on woody heath-gorse vegetation is less well established than for other types. Calibrated with faecal recovery values obtained from previous validation pen studies with animals fed diets of known proportions of the main 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:
3.3 Were all important outcomes assessed? Were all important positive and negative	☐++	Comments:

effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments:
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Only one season. Assumed to be long enough for this type of study.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments: Sample size of faecal and herbage samples not known.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: Effects of different faecal recovery values from validation studies and effects of grazing species, date and interactions on diet composition
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: ANOVA with t-tests for comparison of means for effects as listed above.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: p Values given for all comparisons
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?		Comments: Unreplicated, animals only grazed together, and at one stocking rate. May be effects of

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐-	intra and inter-specific competition.
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: Limited small-scale study.

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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

Study details	Authors	Ferriera, L. M. M., Oliván, M., Rodrigues, M. A. M., García, U. & Osoro, K.
	Year	2005
	Aim of study	To investigate the diet composition of goats and sheep grazing together on gorse-heathland on two grazing dates with different availability of preferred species (perennial rye-grass)
	Study design	2
	Quality score	-
	External validity	-
Population and setting	Source population	Not reported in detail – heatlands in LFA of NW Spain.
	Eligible population	Eligible area is gorse heath with patches of grass and clover. Representativeness not stated, but assumed to be.
	Inclusion and exclusion	Selection not described – likely to be subjective, but chosen to have typical characteristics and vegetation composition. An area of 5 ha sown grassland in the 22 ha

	criteria	plot.
	Setting	Hill experimental farm at 1000m asl in NW Spain
Methods of allocation to intervention/control	Methods of allocation	Single 22ha plot. Likely to have been existing part of the experimental farm.
	Intervention description	Both sheep and goats grazed in same area at same numbers (42 of each) with faeces sampled at two points at start and end July
	Control/comparison description	NA
	Sample sizes	One plot. Number of faecal samples not reported.
	Baseline comparisons	NA
	Study sufficiently powered	No power calculation. No replication. Sample size of faecal and herbage samples not known.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Objective measures of diet composition from alkane concentrations in faeces. Use of the method on woody heath-gorse vegetation is less well established than for other types. Calibrated with faecal recovery values obtained from previous validation pen studies with animals fed diets of known proportions of the main species.
	Secondary outcome measures	
	Follow-up periods	Study over one grazing season – samples just under one month apart
	Methods of analysis	ANOVA with t-tests for comparison of means for effects of different faecal recovery values from validation studies and effects of grazing species, date and interactions on diet composition

Results		The faecal recovery values used in the calculations (from validation studies) were significantly affected by estimates of the proportion of rye grass and heath species in the diet of both sheep and goats. There was no effect on faecal recoveries of the calculated proportions of western gorse, which was always zero. At the first sampling in early July when grass availability was high, there was a significant effect of animal species on composition of diet, with heaths being the greatest proportion of goat diet (68%) and rye-grass for sheep (79%). By the end of July, when grass availability had decreased, there was no significant difference between animal species (61% in goats, 54% in sheep). Sheep are grass grazers when availability is high, and browsers when grass becomes limited. Goats however were shown to prefer heath species even when grass availability is high. The use if different alkene recovery values in the calculations significantly affects the estimates of composition of the diets of both species, despite calculation in controlled metabolic pens using different combinations of the main vegetation types studied in this experiment.
Notes	Limitations identified by author	Use of different alkane faecal recovery values significantly affects the estimates of composition despite being derived in controlled validation study.
	Limitations identified by review team	Mixed grazing – may affect grazing patterns so diet results different from species grazing individually
	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	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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

Study details	Authors	Fisher, G.E.J., Scanlan, S. & Waterhouse, A
	Year	1994
	Aim of study	To investigate the effects of goat and sheep grazing on semi-natural pastures and assess the consequences of this diversification on sites of wildlife conservation value.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Semi-natural hill pasture (rough grazing), but not described
	Eligible population	Likely to be representative of at least some of the common hill pasture communities but again not described in detail
	Inclusion and exclusion criteria	

	Setting	Hill grazing land, Scottish Agricultural College Farm, Crainlarich, west Perthshire, Scotland
Methods of allocation to intervention/control	Methods of allocation	Selection of paddock areas not reported. Grazing treatments not replicated.
	Intervention description	Sheep, goat and mixed treatments. Described in terms of liveweight per ha per yr. Grazed in summer. Some variation between years on treatment liveweight and grazing period (started later in first year)
	Control/comparison description	Three grazing treatments – arguably sheep only is the control.
	Sample sizes	One paddock (2.9 ha) per treatment. Twenty height measurements for each of eight species taken every 10 days. Cover measured in four quadrats per community.
	Baseline comparisons	Not reported in this paper, although baseline vegetation data was recorded and analysed using classification and ordination techniques
	Study sufficiently powered	No analysis, but low powered as treatments not replicated.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation presence and cover in sub-divided quadrats, placed randomly but stratified by community. Height of each main species measured at twenty points every ten days.
	Secondary outcome measures	Liveweights difference at start and end of grazing period.
	Follow-up periods	Three years is about the minimum time to detect grazing related change , and will not detect medium-long term vegetation change
	Methods of analysis	Graphs of mean heights over time of key species in each treatment. No statistical analysis presented. Ordination of vegetation data mentioned but not presented. Repeat vegetation measurements and ordination planned in final year (This is an

		interim paper)
Results		Heights of bog myrtle, rush species and purple moor-grass were lower throughout most of the observation periods in goat and mixed grazing paddocks, although less apparent for rushes in the most recent season (not complete though so may diverge). Grazing on bog myrtle appeared to be delayed until the end of July in sheep and goat mixed compared to goat-only grazing. Some evidence (weak) that sheep graze heather shorter than goats. No indication of differences in grazing on deer grass, *Agrostis- festuca* or mat grass.
Notes	Limitations identified by author	
	Limitations identified by review team	Lack of replication
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	

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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?
Study Citation	Fisher, G.E.J., Scanlan, S. & Waterhouse, A. The ecology of sheep and goat grazing in semi-natural hill pastures in Scotland. Grassland and Society, 286-289 (This appears to be an interim paper – track down final paper?)
Study Design Category	2
Assessed by & when	D Martin 18/11/12

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: Semi-natural hill pasture (rough grazing), but not described
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Likely to be representative of at least some of the common hill pasture communities but again not described in detail
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: Not described, and methods of selection not covered. Likely to be chosen subjectively. No replication. Main plant communities recorded in paddocks are typical of acid grassland and wet heath vegetation.

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: Selection of paddock areas not reported. Grazing treatments not replicated.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Sheep, goat and mixed treatments. Described in terms of liveweight per ha per yr. Grazed in summer. Some variation between years on treatment liveweight and grazing period (started later in first year)
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: Grazed for three years – minimum time for reliable measures of grazing effects?
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐NR	Comments:
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Main communities mentioned are typical. This site in Central Highlands will not fully represent the geographical range and biological variation in the communities recorded.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Sheep grazing treatment will broadly reflect usual practice.

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: Vegetation presence and cover in sub-divided quadrats, placed randomly but stratified by community. Height of each main species measured at twenty points every ten days.
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: Final year measurements planned.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Outcomes measured in terms of heights and species cover/ frequency. Some negative species such as Nardus included
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: Three years is about the minimum time to detect grazing related change , and will not detect medium-long term vegetation change

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: Not reported in this paper, although baseline vegetation data was recorded and analysed using classification and ordination techniques

4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: No analysis, but low powered as treatments not replicated.
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: Graphs of mean heights over time of key species in each treatment. No statistical analysis presented. Ordination of vegetation data mentioned but not presented. Repeat vegetation measurements and ordination planned in final year (This is an interim paper)
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: No analysis presented
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 not replicated. Limited analysis of herbage height data, but this is interim paper (Need to check/ source any final paper)
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: Limited findings, of low reliability, but does provide some pointers of likely grazing preferences/ impacts.

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 impact does grazing two contrasting heathland communities have on diet selection by cattle and sheep?
Study Citation	Fraser, M.D., Theobald, V.J., Griffiths, J.B, Morris, S.M. and Moorby, J.M. (2009) Comparative diet selection by cattle and sheep grazing two contrasting heathland communities. Journal of Agriculture, Ecosystems and Environment 129. 182-192.
Study Design Category	2
Assessed by & when	J Bradley 03/01/13

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 flock and herd of sheep and cattle. 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: Six mature, barren females of each breed/species. Age and status of population not fully representative of the source population. Breeds/species were not fully representative of the source population due to wide range of breeds in source population. Rare breed sheep were not represented.
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐+	Comments: The breeds/species were generally representative of eligible population, rare breed sheep not represented.

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 4 ha plots of upland heath and mire subdivided into 4 sub plots. One plot had high percentage cover of C*alluna vulgaris* (61%), one plot had low cover of *Calluna vulgaris* (8%). The two plots were grazed by Welsh Mountain sheep, Scottish Blackface sheep, Welsh Black cattle and Continental cross cattle. Each breed/species was randomly assigned to a sub plot at the beginning of the measurement period then moved to the next sub plot daily. The grazing was carried out on the low plot over a one week period in both July and September 2004 and to the high plot for one week in July and September 2005. Stocking rates between species not comparable.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Treatments well described with duration so could be replicated. Sites grazed by all breeds/species during each measurement period. Stocking rates between species not comparable but comparison between breeds/species was appropriate.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Measurement period was short (1 week) but all animals had previous experience of grazing hill areas and all were grazed on areas adjacent to the measurement plots for at least 2 weeks prior to the start of each measurement period. Period of rotation of animals around sub plots during measurement period not clear- 4 sub plots/4 different breeds/species grazed over 1 week therefore it would appear that 1 sub plot was not grazed by one of the breeds/species to the same degree. Possible source of bias. Exclusion of Continental x cattle from low site may have introduced some bias, not clear if extra grazing by Welsh black cattle took place to compensate
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: No apparent contamination.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional	☐++	Comments: None apparent.

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: Age and status of population not fully representative of the source population. Breeds/species were not fully representative of the source population due to wide range of breeds in source population. Rare breed sheep were not represented.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The use of sites with high and low percentage cover of *Calluna* did represent usual UK practice whilst not fully covering the full range of *Calluna* cover. Stocking rates were higher than usual practice.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Botanical composition, herbage biomass, diet composition and dietary preferences all measured and reliable. Exclusion of Continental x cattle from low site may have introduced some bias, not clear if grazing regime was amended to compensate.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐-	Comments: No, continental x cattle not measured at low site.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐-	Comments: No, continental x cattle not measured at low site.
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, Botanical composition, herbage biomass, diet composition and dietary preferences all measured and reliable.
3.5 Were there similar post-treatment time intervals in exposure and comparison	☐+	Comments: Yes

groups?		
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Assessed during 4 weeks over two years, sufficient to show some significant results. Longer assessment period may be required to assess long term effects.

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, continental x cattle not measured at low site. No adjustment made. Unclear if adjustments made during measurement period.
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: No power analysis given.
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments: Botanical composition, herbage biomass, diet composition and dietary preferences.
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: One way analysis of variance carried out on first principal component to investigate effect of species on variation. Analysis of variance using Genstat 8.1 used on dietary components which was then subject to angular transformation. Diet selection was quantified using Jacobs (1974) modification of Ivlev's electivity index.
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 .
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of	☐+	Comments: Treatments were implemented well but weak replication. Sources of bias were minimised but welfare issues introduced possible source of bias. No apparent adjustment for potential bias introduced

bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?		by loss of continental x cattle grazing at low site. A good example of design of a controlled grazing experiment.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Due to the wide range of breeds of both sheep and cattle within the source population it would be difficult to extrapolate the results fully but the findings are generalisable.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What impact does grazing two contrasting heathland communities have on diet selection by cattle and sheep?

Study Details	Authors:	Fraser, M.D., Theobald, V.J., Griffiths, J.B, Morris, S.M. and Moorby, J.M.
	Year:	2009
	Aim of study:	To test the effects on diet selection by cattle and sheep grazing two contrasting heathland communities 1. Low percentage cover of *Calluna vulgaris*. 2. High percentage cover of *Calluna vulgaris*
	Study design:	2
	Quality Score	+
	External validity:	+
Population and setting	Source population:	Population is the UK flock and herd of sheep and cattle. Not described in detail.
	Eligible Population:	Six mature, barren females of each breed/species. Not representative of the source population.

	Inclusion & exclusion criteria:	
	Setting:	Upland heath and mire in Ceredigion, Wales. Situated between 510m and 580m a.s.l.
Methods of allocation to intervention / control	Methods of allocation:	Subjective – 1 replicate of each treatment – randomised.
	Intervention description:	Two 4 ha plots of upland heath and mire subdivided into 4 sub plots. One plot had high percentage cover of C*alluna vulgaris* (61%), one plot had low cover of *Calluna vulgaris* (8%). The two plots were grazed by Welsh Mountain sheep, Scottish Blackface sheep, Welsh Black cattle and Continental cross cattle. Each breed/species was randomly assigned to a sub plot at the beginning of the measurement period then moved to the next sub plot daily. The grazing was carried out on the low plot over a one week period in both July and September 2004 and to the high plot for one week in July and September 2005. The Continental cross cattle had to be excluded from the low site study on welfare grounds.
	Control / comparison description:	Comparison of two treatments, but no control as such.
	Sample sizes:	Two treatments , with grazing by six mature, barren females of each of the four different breed/species. Grazing over a one week period in both July and September of 2004 and 2005. Vegetation and livestock measurements taken before and after treatments . Each individual animal was treated as a replicate.

	Baseline comparisons:	Baseline botanical composition and biomass. Diet composition of each breed/species on each treatment..
	Study sufficiently powered	No power analysis given.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Primary outcome measures:	Comparison of diet composition and preferences of each breed/species.
	Secondary outcome measures:	
	Follow-up periods:	No follow up period.
	Methods of analysis:	One way analysis of variance carried out on first principal component to investigate effect of species on variation. Analysis of variance using Genstat 8.1 used on dietary components which was then subject to angular transformation. Diet selection was quantified using Jacobs(1974)modification of Ivlev's electivity index.
Results		There were significant differences in principle components analysis of faecal concentrations between cattle and sheep and between sheep breeds grazing the low site in both July and September (cattle breeds n/a). There were also significant differences between sheep and cattle in July and September on the high site, but significant differences between sheep breeds only in September and between cattle breeds in July. The results indicate generally greater variation in the dietary choices of sheep compared to cattle,

		and significant effects of species and sampling session indicate differences in dietary choices between cattle and sheep, and at different times of the year, respectively. At the low heather cover site in July sheep consumed more heather than cattle, and Scottish blackface consumed more heather than Welsh Mountain sheep, but WM consumed more of other dwarf shrubs. The September diet of both breeds contained less heather, and the WM consumed less other dwarf shrub. At the high heather cover site in July there was no significant difference in the diet of the two sheep breeds, and the only significant cattle breed difference was the continental cross animals consuming more of other dwarf shrub. The sheep diets contained less Nardus and more broad-leaved grasses and heather than cattle. In September the diet of the two cattle breeds remained similar, there were differences between sheep, with the Scottish blackface consuming more heather and cotton grass. Diet preference showed differences between the Welsh Black cattle and Scottish Blackface sheep, and the Welsh Mountain sheep at the low site in both July and September. There were diet preference differences between sheep and cattle at the high site in July and September. The SBF sheep showed greater variation between the two sessions than WM. There were greater differences in the cattle between grazing periods, with decreases in other dwarf shrub, fine grasses and cotton grasses and increases in sedges and broad-leaved grasses, despite the latter being at lower cover than July. The pattern was similar for both breeds. Consumption of *Calluna* was comparatively low on both sites by all breeds/species with grasses being the preferred species making up the bulk of the diet. SBF seemed to have weaker avoidance of heather at the low cover site than the high, where all species were seen to avoid heather. Nardus was selected by cattle at the high heather cover, but only weakly at the low cover site. Selection of fine leaved grasses was high for all species at each sampling period, and selection of brad-leaved grasses by cattle increased in September. The study highlights between-breed differences in sheep and the need for studies of grazing behaviour

		and preference of different sheep breeds to develop more effective management guidelines.. Much modelling and experimental work done on SBF in Scotland. There was greater similarity in cattle breeds, suggesting commercial breeds could deliver desired environmental outcomes. However there was shown to be greater chance of welfare and productivity issues with using such stock in marginal areas. Both cattle and sheep were found to be selective grazers, consuming grass in preference to dwarf shrub. Even at high (60%) heather cover this species formed a low proportion (<10%) of the diet of both species, with the proportion slightly higher for sheep than cattle. The diet of the two cattle breeds, Welsh Black and Continental cross showed very little difference between breeds. There was however greater within species variation in sheep, and Scottish blackface were seen to increase the proportion of heather in their diet in September at the high cover site, unlike Welsh Mountain. The results suggest that better information on grazing preferences of different sheep breeds could help meet different objectives more effectively. It also appears that commercial breeds of cattle could deliver desired environmental outcomes. However there was shown to be greater chance of welfare and productivity issues with using such stock in marginal areas.
Notes	Limitations identified by author:	The sampling sessions may have been too close together to show seasonal variations. Exclusion of continental x cattle from low site study.
	Limitations identified by review team:	Limitations in extrapolating finds from small plot experiments to grazing behaviour on open hill, where vegetation choice is greater. Only mature, barren females used, population more likely to be females with young at foot in study period. Grazing period stated to be a week in which all animal types grazed the measurement area – unclear how each sub plot was grazed equally with 4 animal types rotated on a daily basis. Exclusion of continental x cattle may have changed grazing balance on low site study so introducing bias.

	Evidence gaps and/pr recommendations for further research:	Longer study period/with more representative population/stocking rate of livestock.
	Sources of funding:	Defra, English Nature and Countryside Council for Wales.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	What are the impacts on sward composition and stock performance of grazing *Molinia*-dominant grassland? Are there different effects on sward composition from grazing with cattle of sheep? How does stock performance (liveweight gain) change over time when grazing *Molinia*-dominant grassland?

Study Details	Authors:	(a)Fraser, M.D., Theobald, V.J., Dhanoa, M.S. & Davies, O.D. (b) Fraser, M.D., Theobald, V.J., Vale, J. & Evans, G.
	Year:	2011 (a) 2006 (b)
	Aim of study:	To test the effects on sward composition of long term grazing by cattle and sheep of *Molinia*-dominant grassland. To test the effects on animal performance of summer grazing of *Molinia*-dominant grassland.
	Study design:	2
	Quality Score	(a)+ (b)+
	External validity:	(a)+ (b)+
Population and setting	Source population:	Population is the UK extent of *Molinia*-dominant grassland. Not described in detail.

	Eligible Population:	The experimental sites consisted of 2ha plots of rank *Molinia*-dominant grassland. Not representative of grazed *Molinia*-dominant grassland.
	Inclusion & exclusion criteria:	
	Setting:	*Molinia*-dominant grassland at Pwllpeiron Research Station in Ceredigion, Wales. Situated at approx. 540m a.s.l with annual rainfall of 1800mm.
Methods of allocation to intervention / control	Methods of allocation:	Subjective – 2 replicates of each treatment – not randomised.
	Intervention description:	2 ha plots of *Molinia*-dominant grassland which hasn't been grazed for over 20 years. Three treatments were applied to the plots. 1. No grazing, 2. Summer grazing with cattle, 3. Summer grazing with sheep. The treatments were applied over an eight year period.
	Control / comparison description:	Comparing three treatments, but no control as such.
	Sample sizes:	Two plots of each treatment , with summer grazing by different species (at equivalent stocking rates) being carried out as two treatments. Length of grazing period varied on an annual basis but not between species. Vegetation and livestock measurements taken before and after treatments .
	Baseline comparisons:	Botanical composition and sward heights, *Molinia* utilisation and biomass and livestock weight and condition scores recorded at beginning and end of each treatment.

	Study sufficiently powered	No power analysis given.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Primary outcome measures:	Botanical composition and sward heights, *Molinia* utilisation and biomass and livestock performance.
	Secondary outcome measures:	
	Follow-up periods:	Assessed over eight years but as reported in study significant changes to vegetation composition may be a longer term effect.
	Methods of analysis:	ANOVA was carried out on botanical composition with treatment and year as factors. This was followed by repeated measures of Anova with Greenhouse-Geissers estimate of to modify the degrees of freedom. Animal performance was analysed in a similar way. Meta analysis methods were used when individual estimates of a quantity were combined to obtain an overall estimate.
Results		*Molinia* cover and biomass was reduced during the grazing season by cattle grazing, the increase in cover in the sheep grazed plots was half that of the ungrazed plots whilst biomass was also less in the sheep grazed plots than the ungrazed plots.. Significant time effects were identified with regards to *Molina* (increase), broad-leaved grasses (increase), fine-leaved grasses (decrease) and dwarf shrub (decrease) prior to grazing over the study period. There were significant effects following grazing over the course of the experiment on *Molinia* (increase), *Nardus* (increase) and dwarf shrub (decrease). Type of management had little effect on sward composition with only *Molinia* showing a significant change (increase) both before and after grazing.

		Cattle weights improved across 2002 and 2003 but this performance was not sustained and in 2007 and 2008 the cattle lost weight over the grazing period. The loss of weight was thought to be due to poaching caused by exceptionally wet weather causing contamination of the herbage. Sheep weights were significantly affected by year with changes being consistently positive.
Notes	Limitations identified by author:	Despite treatments being imposed for eight years the length of time for change to sward composition on some sites may take considerably longer. Adverse weather conditions in the final two years of the study may have affected cattle performance. Escape of cattle in 2006 resulted in no performance data for that year. Only heifers were used, mature cattle have been shown to be less selective feeders (Grings et al, 2001) and so may have more impact on sward composition.
	Limitations identified by review team:	Limitations in extrapolating finds from small plot experiments to grazing behaviour on open hill, where vegetation choice is greater. Use of hoggets in 2002-may have added bias. Study site only recorded as having *vaccinium* prior to cessation of grazing 20years earlier, with no seed bank was it reasonable to expect a significant increase in dwarf shrub?
	Evidence gaps and/pr recommendations for further research:	Longer study period/study on grazed *Molinia* at beginning of experiment. Effect of mature cattle grazing on sward composition.
	Sources of funding:	Defra

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 impacts on sward composition and stock performance of grazing *Molinia*-dominant grassland? Are there different affects on the composition of *Molinia*-dominant grassland when grazed by sheep or cattle? What was the affect of grazing *Molinia*-dominant grassland on the performance of cattle and sheep?
Study Citation	Fraser, M.D., Theobald, V.J., Dhanoa, M.S. & Davies, O.D. (2011) Impact on sward composition and stock performance of grazing *Molinia*-dominant grassland. Journal of Agriculture, Ecosystems and Environment 144. 102-106. And Fraser, M.D., Theobald, V.J., Vale, J. & Evans, G. (2006) Effects on animal performance of summer grazing of *Molinia*-dominant semi-natural rough grazing. Journal of Biodiversity Science and Management 2. 247-248.
Study Design Category	2
Assessed by & when	J Bradley 28/11/2012

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.	☐+	Comments: Population is the UK extent of *Molinia*-dominant grassland. 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 were rank *Molinia*-dominant grassland which had not been grazed for over 20 years. Representative of rank *Molinia*-dominant grassland but not grazed *Molinia*-dominant grassland. Only recorded dwarf shrub prior to cessation of grazing-*Vaccinium*. Not representative of degraded upland heather heath.
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 experimental sites were typical of rank *Molinia*-dominant grassland and within altitudinal and climatic 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. Treatments consisted of sites with 1. No grazing, 2. Grazing with cattle, 3. Grazing with sheep. Sites subjected to treatments over same summer periods at comparable stocking rates (lu's) for each species. Treatments repeated for eight years.
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 with tabulated timing and duration of treatments so could be replicated. Sites grazed by same species each year. Annual stocking rates are typical of farm practice but are concentrated in time and space for the purpose of the study. Comparison between species was appropriate.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: Exposure periods were appropriate (49-76 days). Repeat of exposures (8 years) was also adequate. Problems with shortened exposure periods due to movement restrictions (foot and mouth 2001) and adverse weather conditions (2005-2008) reported. Not considered to cause important bias. Due to escape of cattle no cattle performance measurements were recorded in 2006. This is not deemed to cause important bias. Use of hoggets in 2002 may have caused bias.
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 apparent 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: None apparent.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Likely to be representative of rank *Molinia*-dominant grassland but not grazed *Molinia*-dominant grassland. Not representative of degraded upland heather heath.
2.7 Did the intervention(s) or control		Comments: The overall stocking rates calculated on an

comparison(s) reflect the usual UK practice(s)?	☐+	annual basis are broadly in line with practice. A mixed grazing intervention may also have been appropriate.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐++	Comments: Botanical composition measurements taken using random locations. Grouping of plants with similar functional and morphological characteristics – objective. Measurement of plant height recorded with a sward stick – objective. Link between *Molinia* utilisation and biomass measured using exclusion zones within treatment sites – objective. Live weights of cattle and sheep recorded after fasting at beginning and end of each grazing period – 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: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes, Sward height, botanical composition of sward, *Molinia* biomass and utilisation and livestock live weights.
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: Assessed over eight years, sufficient to show some significant changes but other changes may require longer as described in paper.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]?	☐++	Comments: Yes

Were there any differences between groups in important confounders at baseline?		
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments: No power analysis given.
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments: Sward height, botanical composition (plant groups as a %) of sward, *Molinia* biomass and utilisation and livestock liveweights before and after treatment given.
4.4 Were the analytical methods appropriate? Were any important differences in post-treament time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐++	Comments: ANOVA was carried out on botanical composition with treatment and year as factors. This was followed by repeated measures of Anova with Greenhouse-Geissers estimate of to modify the degrees of freedom. Animal performance was analysed in a similar way. Meta analysis methods were used when individual estimates of a quantity were combined to obtain an overall estimate.
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 given for regression equation R^2 values.
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. Sources of bias were minimised well. Aa good example of a controlled grazing experiment.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to	☐+	Comments: Findings are generalisable to specific habitat – rank *Molinia*-dominant grassland and to a lesser extent grazed *Molinia*-dominant grassland – and specific livestock types but as described in study, livestock of different ages may perform differently.

determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?		

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Fryday, A. M.
	Year	2001
	Aim of study	To investigate the effects of grazing on lichen amount and growth in montane vegetation
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Only described in very general terms. No detail of typical vegetation types.
	Eligible population	Three main sites through key examples of the habitat in England, Scotland and Wales, and two other areas in Scottish Highlands included. These areas together should be fairly representative of the geographical range and variation of upland/ montane habitats in general, however Scottish site is low lying and although montane in character due to its latitude, the community has elements of sub-montane vegetation.
	Inclusion and exclusion	Existing enclosures sited on montane vegetation.

	criteria	
	Setting	Moor House NNR, Upper Teesdale (550-840m), n England; Inchnadamph (250m), NW Scotland; Crib Goch (850m), Snowdonia, Wales.
Methods of allocation to intervention/control	Methods of allocation	No indication of how exclosures originally chosen, as the main ones have been in place for some years. Assumed they were chosen as typical of the commonly-occuring vegetation types in the locality, although Beinn Eighe quadrats were targeted at a particular lichen species (Arctic kidney lichen *Nephroma arcticum*). However exclosures at main sites not chosen specifically to investigate lichen populations
	Intervention description	Large herbivore razing exclusion in long-term fenced areas.
	Control/comparison description	Background grazing levels on surrounding hill land.
	Sample sizes	Seven plots at Moor House, one at Snowdonia and one at Inchnadamph (but sampled in two places)
	Baseline comparisons	None- exclusion plots had been in place for in some cases 40+ years before the study
	Study sufficiently powered	No statistical analysis, only basic comparisons inside and out.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Measures of species cover/ frequency and lichen biomass.
	Secondary outcome measures	Vegetation height
	Follow-up periods	Plots in place for different lengths of time – over 40 years in the case of Moor House.

		Grazing regime outside of enclosures may have varied over time.
	Methods of analysis	No statistical analysis, only basic comparisons of vegetation measures inside and out of exclosures.
Results		All sites show greater lichen biomass (particularly fruticose lichens) within exclosures which is particularly pronounced at lower altitudes at the Moor House site (over 17 000 times greater inside at one exclosure and 467 times greater at another). The difference is less at higher altitude and northerly plots. There is evidence of an inverse relationship between lichen diversity and sward height, with more crustose species in short swards, particularly pronounced on limestone areas. Blanket bog plots were an exception to this trend where one exlosure had the same lichen diversity as outside, and another exclosure showed an increase. There is a suggestion from the Welsh site that at higher altitudes grazing does not have such a beneficial effect on species diversity, but this is tenuous. A marked difference was observed in the lichen community inside exclosures over limestone at Moor House compared to outside, and being more similar to inside exclosures on acid grassland. The surface here was shown to be more acidic than outside, suggesting that the build up of vegetation isolates lichens from the effects of the substratum. Removing grazing at low to intermediate altitudes may adversely affect lichen diversity and needs to be carefully considered.
Notes	Limitations identified by author	Only one site is true montane in character (Crib Goch), but this has been compromised by sheep access to the exclosure and it is lichen poor compared to other areas in the study. The observations are largely derived from sub-montane situations
	Limitations identified by review team	Poor study design, lack of statistical robustness.
	Evidence gaps and/pr	More exclosure studies required on lichen-rich vegetation, including some targeted at

	recommendations for further research	restricted species to investigate the role of grazing in their current distribution.
	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	h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Fryday, A. M. (2001). Effects of grazing animals on upland/montane lichen vegetation in Great Britain. Botanical Journal of Scotland, 53, 1-19
Study Design Category	2
Assessed by & when	D Martin 21/11/12

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: Upland/ montane vegetation with lichen component. Only described in very general terms. No detail of typical 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: Three main sites through key examples of the habitat in England, Scotland and Wales, and two other areas in Scottish Highlands included. These areas together should be fairly representative of the geographical range and variation of upland/ montane habitats in general, however Scottish site is low lying and although montane in character due to its latitude, the community has elements of sub-montane vegetation.
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: No indication of how exclosures originally chosen, as the main ones have been in place for some years. Assumed they were chosen as typical of the commonly-occuring vegetation types in the locality, although Beinn Eighe quadrats were targeted at a particular lichen species (Arctic kidney lichen *Nephroma arcticum*). However exclosures at main sites not chosen specifically to investigate lichen populations.

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 treatment is removal of grazing. Initial plot selection not reported. Sample areas from exclosure and grazed area chosen subjectively, but "comparable".
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Basically grazing vs no grazing. No information on levels of grazing outside of exclosures.
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: Snowdon site received some grazing inside the exclosure for some time. However the vegetation was noted as still morel luxuriant inside.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐+	Comments: The nature of the study means that sites will be subject to different environmental, climatic and management regimes, impossible to control for in this type of observational study.
2.5 Is the setting applicable to the UK?	☐++	Comments: Good examples of habitat and geographical range covered including main areas in Wales and N England.

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: Sample area selection was subjective. All species assigned Domin values from within a 4 x 4 m quadrat, except saxicolous lichens which were given a DAFOR value. Estimating cover from one large quadrat is of limited reliability, hence '-' score. Biomass of lichen measured in small 20 x 20 cm quadrats and multiplied up to give a value for the relevé (large quadrat). Increase inside the exclosure calculated by dividing the biomass from inside by that calculated from outside. Soil was sampled in each relevé and pH measured in distilled water using a pH meter.
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: No measurements from outside one of the Inchnadamph exclosures due to deteriorating weather.

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	Comments:
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?	☐+	Comments: Plots in place for different lengths of time. Grazing regime outside of enclosures may have varied over time.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Most for long enough to develop responses to lack of grazing.

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 statistical analysis, only basic comparisons inside and out.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: No attempt made to apportion effects, but postulated that sheep grazing is most likely to be major cause of differences in lichen cover and biomass.
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:
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:

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: Sampling largely subjective and no statistical analysis. Some very large differences in lichen biomass inside and outside exclosures are shown, but varies between sites and altitudes.
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: To an extent, given the large increases in biomass observed: "the overall results are so clear-cut that concerns with the objectivity of the methods are inconsequential". Whilst there are differences inside and out, the methods may magnify these differences. The effect may be generalisable, but possibly not the magnitude.

Evidence Table

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

Study details	Authors	Gardner, Hetherington & Allen
	Year	2002
	Aim of study	1. To quantify the long-term effects of Cambrian Mountains ESA (CMESA) grazing prescriptions (and other reduced stocking densities) on degraded heather and grass dominated moorland plant communities. 2. To assess the viability of the seed bank in relation to heather regeneration on grass-dominated, previously over-grazed moorland plots. 3. To develop a field methodology for quantifying sheep grazing levels on *Nardus* and other grasses. 4. To describe the spatial pattern of grazing in relation to vegetation composition (on grass-dominated and heather/grass moorland) and to determine the levels of seasonal and annual variation in such patterns. 5. To determine the relationship between vegetation change (particularly with regard to the balance of *Calluna, Nardus* and *Vaccinium*) and grazing pressure at the quadrat scale. 6. To develop a spatial model for assessing the impact of variation in grazing pressure on vegetation change in degraded moorland plant communities and to test model

		predictions against field data. To identify appropriate grazing scenarios to enhance the recovery of dwarf-shrub vegetation on degraded moorland
	Study design	2 unreplicated plot experiment
	Quality score	+
	External validity	+
Population and setting	Source population	Not given in detail. Heather and related grass communities in the uplands
	Eligible population	Calluna, Nardus and Vaccinium dominated communities are the subject. Likely to reflect a rangeof upland habitat conditions
	Inclusion and exclusion criteria	Existing experimental plots, established on target vegetation types
	Setting	Pwllpeiron Experimental farm
Methods of allocation to intervention/control	Methods of allocation	Not known, likely to have been subjective to capture target vegetation types
	Intervention description	Expt 1: Cambrian Mountains ESA (CMESA) Tier 1A (1.5 sheep ha^{-1}) and Tier 2A (1 sheep ha^{-1}) stocking rate prescriptions were applied to two paddocks of *Calluna-Nardus* dominated vegetation and two paddocks of *Agrostis-Festuca* dominated vegetation. Sheep are Welsh Mountain. Expt 2: five 1 ha paddocks of *Vaccinium-Nardus* vegetation were established in 1990 and the following five grazing treatments applied to one paddock each: i) CMESA Tier 1A stocking prescriptions applied between April- July, ii) CMESA Tier 2A stocking prescriptions applied between April-October, iii) CMESA Tier 2A stocking prescriptions applied between April- July, iv) CMESA Tier 1A stocking prescriptions applied between April-October

		v) an ungrazed paddock The original ESA prescriptions allowed a stocking rate of 3 sheep per ha, adjusted for an ungrazed period and a 1.9 ewe per ha treatment also applied. Additional 5 paddocks added in 1995 to look at seasonal application of tier 1 (1.5) and 2 (1) stocking rates to previously ungrazed Vaccinium Nardus veg, and a low rate of 0.5. sheep on previously grazed and ungrazed. Provides some replication of Expt 2, but different timescales.
	Control/comparison description	No control as such. ESA rates over six month period may be closest
	Sample sizes	One lot per treatment. Veg samples from 16 1m^2 quadrats
	Baseline comparisons	Vegetation top cover measures in 1995 and DCA used to identify community types at start of this phase
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation composition top cover measures, spatial grazing on Calluna, Nardus and Vaccinium and soil seed bank
	Secondary outcome measures	
	Follow-up periods	Main expts in place 1995-2002

	Methods of analysis	DCA and repeated measures ANOVA, with significance testing. Interactions of year and veg type included.
Results		Four different vegetation types were identified from DCA and fuzzy clustering. These were: i) *Festuca-Agrostis* grassland, dominated by these two genera and also having a high frequency (but low cover) of *Nardus* and *Vaccinium,* ii) *Calluna-Eriophorum* wet heath, characterised by a frequent occurrence of species such as *Erica tetralix, Eriophorum angustifolium* and *E. vaginatum*, iii) *Vaccinium-Nardus* heath, being similar to i) but having a greater cover of *Vaccinium* and *Nardus* and *Calluna-Nardus* heath, a heather dominated group with frequent *Agrostis* sp.*, Festuca* sp.*, Nardus* and *D. Flexuosa* There was little change in top cover of Calluna-Nardus at ESA T1 stocking rate, although there was a decrease in top cover of Calluna in this vegetation under T2 rates. On the whole vegetation remained stable across the plots. Greater annual variation in grass and sedge species was observed within the dwarf shrub vegetation under the Tier 2 stocking rates than under the Tier 1 stocking rate. Although differing significantly in composition, the *Calluna-Nardus* and the *Vaccinium-Nardus* vegetation showed similar changes in the cover of grass and sedge species, *Festuca* and *C. pilulifera* increasing significantly but other species showing little consistent directional change. IN Agrostis-Festuca vegetation A capillaris increased at the expense of F ovina in both treatments, J squarrosus increased in Agrostis-Festuca and Vaccinium-Nardus under tier 1 grazing, probably due to reduced grazing in winter and spring. Accumulation of dead material in both treatments may lead to lower lamb performance over time. In expt 2 there was a significant decline when tier 1 rates were applied, particularly over a 3-month period. The tier 2 rates showed no change in Vaccinium cover when applied over the 3-month period. When the same overall sheep numbers were applied over a 6-month period a

		decline in cover was observed. In the 1995 plots Vaccinium declined across all four treatments and Nardus increased on three Seasonal grazing of *Nardus-Vaccinium* heath led to significant changes in the balance of *Vaccinium* and grasses, particularly *Nardus* and *D. flexuosa,* but had little effect on the regeneration of *Calluna* within this vegetation type. Under most grazing treatments, *Vaccinium* decreased in cover, with corresponding increases in one of these grass species, *Nardus* under grazed conditions and *D. flexuosa* under ungrazed conditions. On the Tir Emrys paddocks, for which the longest run of data were available, *Vaccinium* declined under three of the four grazing treatments. Welch (1998) observed reductions in the cover of *Vaccinium* under grazed and ungrazed conditions on *Calluna-Vaccinium* moorland. The results from this study, albeit from a different vegetation type, tend to support the idea that *V. myrtillus* is sensitive to even relatively low levels of grazing. Cover was only maintained when grazing was concentrated in the three months in spring, which may be because grasses are grazed at this time in preference to vaccinium, which tends to be grazed in Autumn. Results contrast with early predictions of Rushton et al (2002) who postulated that the Tir Emrys plots would move towards heather moorland. There has however been little change in heather cover over the 10 year period. Seed bank studies suggest little Calluna seed under Nardus-Vaccinium and grass dominated communities. Bilberry moorland may be a more viable objective. Spatial variation in grazing. There was a significant difference in the frequency of *Calluna* grazing between the three vegetation types (p=0.005) with higher levels of grazing being recorded at the interface between communities and within the *Calluna* dominated vegetation type than in Nardus. Higher grazing levels on Calluna were recorded at the end of winter than summer. *Vaccinium* and *Nardus* grazing was significantly higher in the interface and *Nardus*-dominated vegetation. In addition, grazing of *Nardus* and *Vaccinium* was higher within these communities during 2001 than in 2000. However, *Nardus* grazing was significantly higher at the end of winter for both of

		these vegetation types whereas *Vaccinium* exhibited different characteristics. Within *Nardus*-dominated veegetation, grazing on *Vaccinium* was greater during the end of summer assessment period than at the end of winter, whereas for the interface community, grazing was greatest at the end of winter. The study has highlighted the importance of spatial distribution, plant productivity and species composition in determining the direction and magnitude of vegetation change on upland dwarf shrub heath. There is evidence that the previous heavy grazing may have reduced productivity and hence competitive ability of Vaccinium, resulting in a lag effect when grazing is reduced.
Notes	Limitations identified by author	
	Limitations identified by review team	Lack of replication, possible limitation of top cover in measuring spread of lower vegetation as height increases.
	Evidence gaps and/pr recommendations for further research	The research has highlighted two gaps in our understanding of the processes driving vegetation change on upland heath. The first relates to a lack of quantitative information on the relationship between plant competition and plant productivity and how this relationship is influenced by grazing animals. The second relates to the role of sub-dominant (minority) species in influencing the direction and level of species change.
	Sources of funding	Defra

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	Gardner, S. M., Hetherington, S. L. & Allen, D. 2002. Assessment of vegetation change and *Calluna/Nardus* interactions in relation to spatial variation in grazing pressure on upland moor. Final Report to Defra/ WOAD Contract BD1211
Study Design Category	2
Assessed by & when	D Martin 25/01/13

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	☐+	Comments: Not given in detail. Heather and related grass communities in the uplands
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: Calluna, Nardus and Vaccinium dominated communities are the subject. Likely to reflect a rangeof upland habitat conditions
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: Existing experimental plots, established on target vegetation types

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: Unreplicated, selected subjectively?
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Expt 1: Cambrian Mountains ESA (CMESA) Tier 1A (1.5 sheep ha^{-1}) and Tier 2A (1 sheep ha^{-1}) stocking rate prescriptions were applied to two paddocks of *Calluna-Nardus* dominated vegetation and two paddocks of *Agrostis-Festuca* dominated vegetation. Sheep are Welsh Mountain. Expt 2: five 1 ha paddocks of *Vaccinium-Nardus* vegetation were established in 1990 and the following five grazing treatments applied to one paddock each: i) CMESA Tier 1A stocking prescriptions applied between April- July, ii) CMESA Tier 2A stocking prescriptions applied between April-October, iii) CMESA Tier 2A stocking prescriptions applied between April- July, iv) CMESA Tier 1A stocking prescriptions applied between April-October v) an ungrazed paddock The original ESA prescriptions allowed a stocking rate of 3 sheep per ha, adjusted for an ungrazed period and a 1.9 ewe per ha treatment also applied. Additional 5 paddocks added in 1995 to look at seasonal application of tier 1 (1.5) and 2 (1) stocking rates to previously ungrazed Vaccinium Nardus veg, and a low rate of 0.5. sheep on previously grazed and ungrazed. Provides some replication of Expt 2, but different timescales.
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: Plots originally set up in 1990, but Ag-Fe plots added in 1995. This phase extended the project from 1998 to 2002. New Vm-Ns plots established in 1995

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: Starting points of plots differ in terms of grazing history. May not be taken account of adequatley
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
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐++	Comments: Likely to be broadly representative of upland heath mosaics on mineral and thin peat soils
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Expt 1 treatments reflect ESA, but not typical farming practice.

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: Species change in permanent quadrats in each plot (first hit/ top cover). Seed bank germination trials from soils in each plot. Grazing pattern of sheep on different species recorded by assessing grazed shoots in fixed plots twice per year.
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?	☐++	Comments:

If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?		
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Some variation between experiments in length of exposure
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Treatments in place for at least 6 years

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: Would have been different with different grazing history, but DCA carried out on the 1995 data to identify different vegetation groups at start of this study period, and included in ANOVA as a fixed factor.
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?	☐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: DCA and repeated measures ANOVA, with significance testing. Interactions of year and veg type included.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for	☐++	Comments:

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 any significant flaws in the study design?	☐+	Comments: Not replicated. Plot areas have different grazing histories. Different veg types accounted for in analysis
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++	Comments:

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	MH Garnett, P Ineson & AC Stevenson
	Year	2000
	Aim of study	Effects of burning and grazing on carbon sequestration in a Pennine blanket bog, UK.
	Study design	1
	Quality score	=QA 5.1 Differences in the amount of C above the SCP 'take-off' will have occurred as a result both of treatments and natural variability in peat accumulation (evidenced by total peat depth varying between 1m & 2m). However, there were only small differences in peat depth within the same block under different treatment plots. +
	External validity	=QA 5.2'Burning is practiced regularly on large areas of moorland in upland Britain to provide uneven-staged stands of heather'
Population and setting	Source population	Hard Hill, part of Moorhouse NNR in the North Pennines AONB. Blanket bog.
	Eligible population	Calluneto-Eriophoretum on blanket bog 1-2m thick Altitude 600-630m Mean annual rainfall 1900mm

	Inclusion and exclusion criteria	
	Setting	Hard Hill, part of Moorhouse NNR in the North Pennines AONB. Calluneto-Eriophoretum on blanket bog 1-2m thick Altitude 600-630m Mean annual rainfall 1900mm
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	Experimental design factorial – 3 different burning treatments (every 10 years, every 20 years, not burnt) x 2 grazing treatments (grazed and ungrazed).
	Control/comparison description	
	Sample sizes	Not stated Details given of experimental layout, sampling procedures and determination of carbon content. Numbers of samples and area of blocks not stated
	Baseline comparisons	Entire area burnt prior to the construction of the experimental plots in1954. The method of burning used is similar to traditional moorland burning
	Study sufficiently powered	An analysis of variance was undertaken using MINITAB version 10.2
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Effect of sheep grazing on carbon accumulation on blanket bog. Measurement of mean mass of carbon above the SCP level Effect of burning on carbon accumulation in blanket bog.
	Secondary outcome measures	
	Follow-up periods	

	Methods of analysis	
Results		'After over 30 years of different management there was no detectable difference in the carbon accumulated under the separate (grazing) treatments' 'Significantly less C was contained above the SCP level under the treatment which had been burnt every 10 years compared with the unburnt treatment, implying that this management practice contributes to carbon dioxide emissions through i) decreasing the rate of peat accumulation, ii)stopping peat accumulation and/or iii) reducing C stores by burning existing surface peat. It is not possible to establish which of these processes dominated at this site.'
Notes	Limitations identified by author	Density of sheep was very low – 0.2-2 sheep/ha Date of SCP level uncertain so actual rate of C accumulation not determined
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	EN, University of Newcastle, Institute of Terrestrial Ecology, Department of Environment

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	Effects of burning and grazing on carbon sequestration in a Pennine blanket bog, UK. MH Garnett, P Ineson & AC Stevenson The Holocene 10,6 (2000), pp 729-736
Study Design Category	1
Assessed by & when	Alison Hiles 20/2/2013

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: Hard Hill, part of Moorhouse NNR in the North Pennines AONB. Blanket bog.
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: Calluneto-Eriophoretum on blanket bog 1-2m thick Altitude 600-630m Mean annual rainfall 1900mm
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: Experimental design factorial – 3 different burning treatments (every 10 years, every 20 years, not burnt) x 2 grazing treatments (grazed and ungrazed). Entire area burnt prior to the construction of the experimental plots in1954. The method of burning used is similar to traditional moorland burning

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: Each of the 6 treatment plots is replicated in a random pattern in 4 blocks all located on a uniform and generally uneroded gentle slope to the south east of the summit of Hard Hill.
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: Details given of experimental layout, sampling procedures and determination of carbon content. Numbers of samples and area of blocks not stated
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 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:
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: Detailed analytical methods described for spheroidal carbonaceous particle (SCP) determination. The reliability of the profiles of SCP was tested in 1998 by Rhodes and found that the vast majority of particles were clearly visible and easily counted. Profiles of charcoal concentration were also measured and provided chronological information supporting the SCP records.
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: Entire area burnt prior to the construction of the experimental plots in1954. The method of burning used is similar to traditional moorland burning
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: An analysis of variance was undertaken using MINITAB version 10.2

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: Differences in the amount of C above the SCP 'take-off' will have occurred as a result both of treatments and natural variability in peat accumulation (evidenced by total peat depth varying between 1m & 2m). However, there were only small differences in peat depth within the same block under different treatment plots.
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?	✔ ☐++ ☐+ DM ☐-	Comments: Differences in the amount of C above the SCP 'take-off' will have occurred as a result both of treatments and natural variability in peat accumulation (evidenced by total peat depth varying between 1m & 2m). However, there were only small differences in peat depth within the same block under different treatment plots. Only grazed vs ungrazed
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: 'Burning is practiced regularly on large areas of moorland in upland Britain to provide uneven-staged stands of heather'

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	C Gordoni, BA Emmetti, MLM Jones, T Barden, J Wildig, DL Williams, C Woods, SA Belli, B Pugh, DA Norris, TW Ashenden, SP Rushton and RA Sanderson
	Year	2001
	Aim of study	Aims in relation to grazing aspects are to determine the interaction between grazing pressure and nitrogen deposition on the re−establishment of dwarf shrubs and species richness.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	upland moorland dwarf shrub heath and acid grassland communities. Typical of upland sheep-grazed open hill, including degraded heathland under restoration. The resource is not described in detail in this study.

	Eligible population	The experiment utilises long-term experimental areas at Pwllpeiran experimental hill farm, mid-Wales. Well described in terms of altitude, soils, rainfall and vegetation types (U5 grassland derived from historically overgrazed heather dominated heath)
	Inclusion and exclusion criteria	
	Setting	Pen y Garn Hill at ADAS Pwllpeiron experimental farm, mid-Wales. Study site at an altitude of 600m.
Methods of allocation to intervention/control	Methods of allocation	Continuation of two pre-existing grazing treatments. No replication. Original allocation of paddocks likely to have been largely subjective.
	Intervention description	In final paper the actual grazing livestock numbers in each year are given. Basis if the stocking rates given in earlier report (96/20) as 700 and 1400 grazing for light and heavy grazing treatments respectively. The heavy treatment had been adjusted from the original by increasing the ESA rate by 30%, but applying over 6 months. This standardised the grazing period in the two treatments. N addition in 3 small blocks within paddocks. Method and frequency of application described. Measurements also taken from ungrazed and 'moderately' grazed comparisons, as continuations from previous study.
	Control/comparison description	
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	NO power analysis given. Grazing un-replicated.

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		Monitoring of soil water N leaching indicates that soils are already N saturated. Grazing pressure was not found to influence rates of N-cycling or N losses under ambient or elevated deposition rates, with differences found only across grazed/ ungrazed boundaries. It is suggested the wetness of the site and compaction may have had an effect – providing anaerobic conditions for denitrification and loss of N to atmosphere. Sensitivity of species such as mosses and lichens and bilberry to N deposition appeared to be highest at the lower grazing pressure possibly due to increased structure for capture, increased competition for light, or lower phosphate limitation. The latter hypothesis was tested through an application of P to one set of treatment plots in the final year. No effect on species cover or production was observed suggesting P limitation is not contributing to the differential grazing response. A mesocosm experiment involving defoliation as well as N additions concurred with the field experiment findings of grazing pressure effect. It is concluded that low grazing allows for a greater proportion of N-sensitive species, as well as increased competition for light. There was some evidence for preferential grazing of N application plots by sheep, but not significant. Thus areas of high n deposition may be doubly stressed through increased grazing and dunging.

		Heavy grazing was found to increase the biomass of mat grass and moss species, and decrease the biomass of bilberry. There was no evidence from the field study for an effect on the re-establishment of dwarf shrub, possibly due to the slow response of these species. There was however an increased growth effect on bilberry in the mesocosm study. In the N plot vegetation studies there were little difference in total biomass between light and heavy grazing, but different relative contribution of species groups. Bilberry and fine grasses fared best under light or no grazing, with more lichens in ungrazed areas and mosses and heath rush in heavily grazed. The increased sensitivity of some species at lower grazing pressures may offset benefits of reduced grazing in areas of high N deposition.
Notes	Limitations identified by author	Previous experiments may mean that many sensitive species already lost from heavily grazed paddock. There was evidence after 4 years that response as increasing here.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	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	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	C Gordoni, BA Emmetti, MLM Jones, T Barden, J Wildig, DL Williams, C Woods, SA Belli, B Pugh, DA Norris, TW Ashenden, SP Rushton and RA Sanderson (2001). Grazing and Nitrogen interactions in upland acid moorland. CCW Research Report 01/22 and preceding reports
Study Design Category	2
Assessed by & when	D Martin 30/10/12

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: upland moorland dwarf shrub heath and acid grassland communities. Typical of upland sheep-grazed open hill, including degraded heathland under restoration. The resource is not described in detail in this study.
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 experiment utilises long-term experimental areas at Pwllpeiran experimental hill farm, mid-Wales. Well described in terms of altitude, soils, rainfall and vegetation types (U5 grassland derived from historically overgrazed heather dominated heath)
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: Five 2ha paddocks previously established to investigate the timing and duration of grazing effects. Likely to have been chosen subjectively, but seen as typical of surrounding vegetation. No replication. This experiment is based on two paddocks that have been subject to 'light' (ESA stocking rate minus 30% over 6 months) and 'heavy' (ESA stocking rate over 3 month) treatments. Vegetation trajectories are different in these two paddocks, reflecting the different grazing impacts.

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: Continuation of two pre-existing grazing treatments. No replication. Original allocation of paddocks likely to have been largely subjective.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: In final paper the actual grazing livestock numbers in each year are given. Basis if the stocking rates given in earlier report (96/20) as 700 and 1400 grazing for light and heavy grazing treatments respectively. The heavy treatment had been adjusted from the original by increasing the ESA rate by 30%, but applying over 6 months. This standardised the grazing period in the two treatments. N addition in 3 small blocks within paddocks. Method and frequency of application described. Measurements also taken from ungrazed and 'moderately' grazed comparisons, as continuations from previous study.
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 five years, 1996-2000. This should be sufficient to detect grazing effects. Some N deposition effects on moss cover apparent within one year of treatment. N application treatments commenced 1997.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: No, differential grazing maintained throughout the experiment.
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?	☐++	Comments: No other interventions other than the main experimental treatment reported.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Site is in Wales, but communities and farming system reasonably comparable to other mid-altitude moorlands in England and UK.

2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Grazing treatments based on original ESA rates, but the light treatment, grazed in summer only is probably below most typical farms, especially as only grazed seasonally, but reflects conservation regimes.

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: Vegetation composition and biomass – dry weight of harvested material. N availability and cycling - soil mineralisation from soil cores, soil water N (nitrate, ammonium), pH. All objectively measured. Measurements made inside and outside of treatment areas for comaparison. In application studies Veg composition, biomass, annual production, nutrient content, soil water chemistry, gaseous fluxes and transformations measured in four replicates of three treatments.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes- in terms of the review the main outcomes are the effects on vegetation, and grazing effects on N leaching.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Yes, Two main treatments continuation from previous experiment but modified at start of new experiment in 1996. Ungrazed and moderately grazed continued from previous work. N application all commenced at same time.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Grazing treatments have been in place for several years, post N treatment intervals long enough to detect change in some of the vegetation parameters.

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: Paddocks similar at baseline, but on different trajectories due to pre-existing grazing treatments. Experiment with two of these grazing states and comparing their interaction with N deposition. However the previous grazing related vegetation change grazing impacts may be a confounding factor. Some evidence that this has affected response of N 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?	☐-	Comments: NO power analysis given. Grazing un-replicated.
4.3 Were the estimates of effect size given or calculable?	☐+	Comments: Yes – differences in mean values presented graphically as bar charts, line graphs etc where data allows. For some effects only direction of change is summarised.
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: Mean values calculated for different measures under different treatments and presented as bar graphs with SE or pairwise means compared between different treatments with indication of significance. Test not given – t-test? N study plot measures tested by ANOVA. Point made that grazing is not replicated, reducing effectiveness in testing for grazing effects. It is assumed that paddock effects largely down to grazing, but other environmental factors will be at play.
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 from t-tests and anova where appropriate, including where results are close to significant (P< 0.05)
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	☐-	Comments: Measurements and analysis are adequate, but possible confounding of previous grazing treatments and resulting differences at baseline.

design?		
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Largely – vegetation types are broadly representative of mid-hill upland habitats, and stocking rates broadly transferable. Animal behaviour in 2ha plots may be different from ranging behaviour over open hill.

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	Grant, S. A., Bolton, G. R. & Torvell, L.
	Year	1985
	Aim of study	To investigate the effects of controlled grazing on blanket bog, for the integration of conservation and improved sheep grazing
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Blanket bog vegetation. Only key species of ericoids and graminoids given
	Eligible population	Plant communities studied were all variants of *Trichophorum – Eriophorum* bog (McVean & Ratcliffe 1962). The site appears typical of blanket bog vegetation
	Inclusion and exclusion criteria	Three fenced areas of 0.1 ha all chosen to have within-site uniformity. Two sites had been burned two years prior to start but a range of species was recovering. The unburned site had fewer species.

	Setting	Lephimore field station, Cowal, Argyle, Scotland. Site on deep peat at 244m
Methods of allocation to intervention/control	Methods of allocation	Non-random. Although each treatment imposed at each of three sites, they are not considered replicates.
	Intervention description	Three treatments, The two year-round grazing systems were grazed other than for four weeks in spring and ten weeks in Oct-Dec. The off-wintered plots were not grazed from Dec-April. Stocking rates presented in sheep grazing days per ha and also as annual averages. Achieved by grazing 3-4 sheep at monthly intervals fro 1-3 days. Sheep rotated round plots at a site and held in adjacent holding paddock for one week prior to each grazing period.
	Control/comparison description	The low stocking rates reported as approximating to the range found for traditional farming systems.
	Sample sizes	Three stocking rates imposed at each site (one off-wintered, two year-round but recently burned vs older). Biomass sampled at 10 quadrats per treatment plot and composition from 20 pins at 20 locations.
	Baseline comparisons	Broadly similar at baseline- initial floristics presented. Sites broadly similar, some differences due to time since burning. Blocks chosen to have good within-site uniformity across treatment plots. Biomass was shown to be similar in the recently burned sites at the start.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each	Primary outcome measures	Above ground biomass and green shoot biomass and change over time. Percentage cover of species

outcome and significance)	Secondary outcome measures	Sward density from multiple contacts per pin.
	Follow-up periods	Studies in place for 11 years
	Methods of analysis	Mainly analysis of variance. No within site replication so pooled residual variance calculated from samples within plots and used to test for effects of site, stocking rate and interactions. This was recognised as underestimating true error and overestimate statistical significance
Results		Biomass increased at all sites over time as dwarf shrub aged, and differences between sites diminished over time. The effect of stocking rate however increased with time, with biomass on the heavy grazed plots (equivalent to 2.22 sheep ha^{-1} annual average) significantly less than on the light and intermediate treatments. Differences in green shoot biomass were smaller and subject to seasonal climatic effects. This was adjusted for by expressing all treatment measurements as a percentage of the measurements from light grazed plots in the site. There was no indication of differences between the low and intermediate grazing levels in green biomass, but the effect of heavy grazing increased with time ($p<0.001$). The pattern held true when adjustments were made for biomass removed by grazing (based on other work), i.e. to give overall productivity. Treatment effects on *C vulgaris* cover was highly significant ($p<0.001$). There was a significant interaction with site cover decreasing with the heaviest stocking rate at the older heather site, whereas there was an increase over time for all grazing pressures at the other sites. Percentage cover of *E tetralix* was significantly higher in the lower two grazing pressures, and there was again a significant interaction with the effect more marked on the older plot. There were similar highly significant differences in cover of *E vaginatum* with stocking rate, with the effect least at the off-wintered site and greatest at the older heather site. Loss of vascular plant cover was most marked in the older plot especially at highest grazing levels and least marked in year-round grazed recently burned plot.

		Grazing intensity was shown to negatively affect sward density at each site. The effect was most marked on the older sward plot ($p<0.001$) and least in the over-wintered plot. There appeared to be increased sensitivity to grazing overall after eight years, which may be related to climatic conditions and resulting reductions in productivity, possibly allied to increasing heather age. Summary: On blanket bog vegetation (approximating to M17 *Scirpus cespitosus-Eriophorum vaginatum* mire) biomass increased over the ten-year survey period as heather aged, and differences between older and recently burned heather diminished. Over time biomass and green shoot production was reduced on heavily grazed (equivalent to 2.22 sheep ha^{-1} annual average) plots compared with light (0.4 sheep ha^{-1} including off-wintering) and intermediate grazing. The effect of heavy grazing on heather cover was much more marked in older heather. Cover of *E vaginatum* was reduced at high stocking rates on year-round systems. Area of bare ground was higher on heavy grazed treatments, and significantly increased over time on the older heather plot. Decrease in sward density was similarly highest in the heavily grazed older heather plot. Overall the sensitivity of vegetation to grazing was greatly influenced by initial composition and age since burning. Sensitivity appeared to increase after eight years, which may be related to climate and increasing heather age.
Notes	Limitations identified by author	No true replication, grazing confined to blanket bog so no access to other grassland types during grazing periods
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	More information on nutritive value and limits to utilization for different bog species, to help define management regimes for different species compositions. Need for longer term studies on grazing effects on indigenous vegetation.
	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	Grant, S. A., Bolton, G. R. & Torvell, L. (1985a). The responses of blanket bog vegetation to controlled grazing by hill sheep. Journal of Applied Ecology, 22, 739-751
Study Design Category	2
Assessed by & when	D Martin 21/12/12

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: Blanket bog vegetation. Only key species of ericoids and graminoids given
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: Plant communities studied were all variants of *Trichophorum – Eriophorum* bog (McVean & Ratcliffe 1962). The site appears typical of blanket bog vegetation
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: Three fenced areas of 0.1 ha all chosen to have within-site uniformity. Two sites had been burned two years prior to start but a range of species was recovering. The unburned site had fewer species.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐-	Comments: Non-random. Although each treatment imposed at each of three sites, they are not considered replicates.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Three treatments, The two year-round grazing systems (one on recently burned, one on older vegetation) were grazed other than for four weeks in spring and ten weeks in Oct-Dec. The off-wintered plots were not grazed from Dec-April. Stocking rates presented in sheep grazing days per ha and also as annual averages. Achieved by grazing 3-4 sheep at monthly intervals fro 1-3 days. Sheep rotated round plots at a site and held in adjacent holding paddock for one week prior to each grazing period.
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 11 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?	☐NR	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Appears to be an example of the habitat in good condition compared to many English sites – reported as having initially high cover of Sphagnum and E tetralix

2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The low stocking rates reported as approximating to the range found for traditional farming systems. However the implementation in small paddocks (put and take for short periods) is dissimilar to ranging livestock.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: Above ground biomass recorded at start and thee-yearly intervals from ten random small quadrats. Floristic composition in each July from pin frame quadrats, twenty groups of 4 x 5 point quadrats. Restricted random sampling approach, with four observers each making a quarter of the observations in each plot. Mutiple hits of spp also recorded as indication of density.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments:
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments:
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Treatments in place for 11 years

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they	☐++	Comments: Broadly similar at baseline- initial floristics presented. Sites broadly similar, some differences due

adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?		to time since burning. Blocks chosen to have good within-site uniformity across treatment 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 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?	☐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: Mainly analysis of variance. No within site replication so pooled residual variance calculated from samples within plots and used to test for effects of site, stocking rate and interactions. This was recognised as underestimating true error and overestimate statistical significance
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 given for statistical differences from 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: No replication, but same treatments imposed at three sites. Long-term study
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?	☐+	Comments:

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

Evidence Table

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

Study details	Authors	S. A. Grant, D. E. Suckling, H. K. Smith, L. Torvell, T. D. A. Forbes and J. Hodgson
	Year	1985
	Aim of study	Comparative study of diet selection by sheep and cattle: the hill grasslands
	Study design	Meta-analyses, systematic reviews of RCTs or RCTs (including cluster RCTs)
	Quality score	=QA 5.1: Data from each site were analysed separately and tested for differences between periods and between days within a period. No significant flaws seen. +
	External validity	=QA 5.2 The different types of sward are sufficiently described – including some of the forbs -and the findings detailed enough for reasonable generalisation of the findings nationally.
Population and setting	Source population	Upland grassland in southern Scotland
	Eligible population	Agrostis/Festuca, Nardus and Molinia sites
	Inclusion and exclusion criteria	Correct grassland type. Altitude 240-280m
	Setting	Cleish Hills, Fife Forest District and Bell Hill, Wauchope Forest, Roxburgh
Methods of allocation	Methods of allocation	The same group of animals was used throughout and the results were accumulated

to intervention/control		over three years from a random sequence of sites across seasons and years
	Intervention description	NA
	Control/comparison description	
	Sample sizes	3 sites, each fenced into 2 adjacent plots each approx 3ha
	Baseline comparisons	Between grazing periods extra grazing by non-experimental animals was provided as necessary to maintain herbage usage at typical levels (c.20-35%)
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Samples of diet selected were collected on days 2,4 and 6 of grazing on the measurement plots, from oesophageally-fistulated animals. One half retained for measurement of in vitro digestibility and the other half to record the botanical composition. Aerial biomass was determined by cutting at ground level. Botanical analysis was done by random point quadrats
	Secondary outcome measures	
	Follow-up periods	Grazing periods in the Cleish Hills were two in each of 1978, 1979 & 1980. On the Molinia at Bell Hill, 2 each in 1979 & 80
	Methods of analysis	
Results		Sheep differed from cattle in 3 main ways: (i) sheep showed greater variability in diet composition both between and within individual animals (ii) sheep but not cattle were able to increase the proportion of certain components in their diet compared with the proportion in the sward, even when the components grew low in the profile or grew in a fine admixture with other components (iii) sheep but not cattle tended to reduce the proportion of certain tall components in their diets compared with proportions in

		the sward. AGROSTIS/FESTUCA SITE – Both broad- and fine-leaved grass leaf the proportions in the diets of sheep and cattle generally suggested neutral selection by both animal species. Sheep avoided grazing grass flower stems and were more efficient in avoiding dead material in the sward. NARDUS SITE – Both sheep and cattle preferentially grazed between-tussock vegetation. Nardus was avoided by sheep and, to a lesser extent, cattle. Again sheep avoided grass flower stems to a great degree MOLINIA SITE -Sheep and cattle diets were most similar in June and became progressively less similar with advance in season. Initially the decline in similarity reflected the difference in the proportions of grass flower stems (low in sheep, high in cattle) but later the high proportion of Juncus spp in the diets of cattle was a major factor
Notes	Limitations identified by author	It was not possible to conclude whether cattle had no preference for forb species or whether their reduced selection ability prevented forb ingestion. This also applies to the intake of dead grass which was much greater for cattle than sheep.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	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	Moorland Gazing and stocking rates
Study Citation	Comparative studies of diet selection by sheep and cattle: the hill grasslands S. A. Grant, D. E. Suckling, H. K. Smith, L. Torvell, T. D. A. Forbes and J. Hodgson
Study Design Category	2
Assessed by & when	Alison Hiles 29/1/2013

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: Agrostis/Festuca, Nardus and Molinia grassland communities in Southern Scotland
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: A table is included stating location, grid reference, altitude, geology, soils and grazing periods. Pre-treatment of the sites is detailed:- All sites undergrazed with considerable accumulation of dead herbage. Agrostis/Festuca site received 1T/ha lime, Molinia site burnt. No treatment for Nardus/Festuca/Deschampsia. Each fenced in 2x~3ha

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: Same group of animals used throughout and results accumulated over 3 years over a random sequence of sites across seasons and years. This avoids confounding animal and site effects but does involve some confounding of season and year differences. Consequences are regarded as unimportant because observations on diet composition are directly related to observations on sward composition and structure.
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: All interventions and measurements were described in minute and exact detail
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: All in southern Scotland but sward types reasonably typical of upland England also.
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: All measurements were objective. Kulczynski' similarity coefficient was calculated for overall comparison of sheep and cattle diets and also for the comparison of sheep or cattle diets with sward composition. Broad comparisons of the similarity coefficient values between periods and over sites were found to be valid
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:
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: As 3.1 Kulczynski' similarity coefficient was calculated for overall comparison of sheep and cattle diets and also for the comparison of sheep or cattle diets with sward composition. Broad comparisons of the similarity coefficient values between periods and over sites were found to be valid
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: Comparisons were made between the intake of fistulated and non-fistulated animals to adjust for potential confounders
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?	✔ ☐++ ☐+ DM ☐-	Comments: Data from each site were analysed separately and tested for differences between periods and between days within a period. No significant flaws seen.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	✔ ☐++ ☐+ ☐-	Comments: The different types of sward are sufficiently described – including some of the forbs - and the findings detailed enough for reasonable generalisation of the findings nationally.

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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

Study details	Authors	Grant, S.A., Torvell, L., Smith, H. K., Suckling, D. E., Forbes, T. D. A. & Hodgson, J.
	Year	1987
	Aim of study	To investigate diet selection and nutrient intake of sheep and cattle grazing together on two dwarf shrub communities (heath and bog)
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Upland heath and blanket bog
	Eligible population	Chosen to reflect common vegetation types on deep peat and peaty podzols. Likely to be fairly typical especially of N Scotland, but only one site for each habitat.
	Inclusion and exclusion criteria	Site selection not described. Likely to be opportunistic e.g. on experimental farms. Only one site per habitat. Plot areas probably chosen subjectively

	Setting	Blanket bog site at Lephinmore, Argyll and heath at Glensaugh, East Grampians. Both sites between 240m and 290m.
Methods of allocation to intervention/control	Methods of allocation	One treatment per site, no replication. There was a degree of randomisation in combinations of livestock individuals and grazing period across these and the sites reported in Grant (1985).
	Intervention description	Each 3h plot grazed with 11 barren suckler cows (blue-grey and Hereford x Fresian) and 10-13 barren blackface ewes. Measurement plots grazed for 6 days in each period following six days on adjacent "run-in" plot. Each plot subject to 4 (BB) or 5(heath) grazing periods at different times of year.
	Control/comparison description	N/A
	Sample sizes	Diet sampled from 3-4 animals of each species on three days of each grazing period. Biomass sampled from fifteen 40 x 20 cm quadrats in the blanket bog plot and six to twelve quadrats in two different ages in the heath plot. Botanical composition and structure assessed at each grazing period from 50 pin hits at 16 to 30 locations in each plot.
	Baseline comparisons	N/A
	Study sufficiently powered	N/R
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Makeup of diet in each period for each species; sward composition at different grazing periods
	Secondary outcome	

	measures	
	Follow-up periods	All measurements made during grazing periods. Both sites grazed for similar number of periods and same duration, although no grazing in 1979 at blanket bog site to let it recover from high utilisation of previous periods.
	Methods of analysis	Analysis of variance on transformed data used to analyse differences in diet components between periods and to measure variation between and within animals of the same species. Anova and Cochrane's t-test used to compare sheep and cattle diet, where within and between animal variation was similar (anova) or different (t-test). Kulczynski's similarity co-efficient used to compare sheep with cattle diets and sheep and cattle diets with sward composition.
Results		Sheep diets were shown to be slightly more variable than cattle diets: on blanket bog between animal variation was similar for sheep and cattle while within-animal variation was greater for sheep; on the heather moor between animal variation was greater for sheep than cattle whereas within-animal variation was similar for both animal species. Whilst sheep and cattle differ significantly for most components of diet in at least some grazing periods on both habitats, on the heath they were similar throughout for proportion of heather leaf, and *Juncus* species (mainly *J squarrosus*). On the blanket bog diet composition is greatly influenced by time of year. There was a lower proportion of dead material in sheep diet throughout the year. The proportion of dead material in cattle diet exceeded the proportion on the sward throughout the season, compared with August – October for sheep. *Eriophorum* spp was lower in sheep for all periods other than April when intake of floral parts was high. Intake of *Molinia*, *Trichophorum* and other grasses and sedges was higher in sheep than cattle in July and September periods. The proportion of these species in sheep diet exceeded their proportions in the sward. Heather intake increased in sheep in October, the only time it exceeded proportion in the sward, but not in cattle. Overall cattle diet was more similar to proportion in the sward throughout

		the year than sheep. Sheep diet was most similar to the sward in October when heather intake was greatest, whilst cattle diet was least similar in this period. In the heath community diet of both species was influenced by time of year. *Calluna* shoots contributed the greatest proportion of sheep and cattle diet in April-May and October –November, with very low proportions in July when both species grazed *J squarrosus*, and sheep also grazed other grasses and sedges. The proportion of *Vaccinium* in diets increase in sheep in May, and both species in July. Both these species are grazed at higher proportions than in the sward throughout spring and summer, whilst heather is present in both diets at much lower proportions in the sward from May-September. Diet similarity to sward was lowest in July for sheep, and the pattern was similar in cattle although was marginally higher than for sheep throughout. Summary: Patterns of diet selection in sheep and cattle were similar when grazing both blanket bog and heather moor, with time of year having a marked effect on selection. This is mainly due to the low preference for heath species and availability preferred graminoids. On blanket bog species selected by sheep tended to have low cover. Cattle were less effective at selection and tended to graze higher proportions of cotton grass leaves. Heather increased in cattle diet in spring when there was much dead cotton grass, and in sheep in October when preferred species had died back. On the heath cattle selected similar species to sheep, but these tended to form patches (Vaccinium and Juncus), with sheep better able to select more scattered palatable grasses. There is evidence that cattle are more reluctant to graze heather than sheep, however they were shown to remove a greater proportion of the woody growth than sheep.
Notes	Limitations identified by author	
	Limitations identified by review team	No replication of sites, lack of grazing in one season on blanket bog.

	Evidence gaps and/pr recommendations for further research	More work needed on interactions between burning and grazing in both these habitats, and with various ratios of dominant species, to assist with management decisions.
	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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?
Study Citation	Grant, S.A., Torvell, L., Smith, H. K., Suckling, D. E., Forbes, T. D. A. & Hodgson, J. (1987) Comparative studies of diet selection by sheep and cattle: blanket bog and heather moor. Journal of Ecology 75, 947-960.
Study Design Category	2
Assessed by & when	D Martin 30/12/12

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: Only very broadly indicated as blanket bog and species poor heather moorland
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: Chosen to reflect common vegetation types on deep peat and peaty podzols. Likely to be fairly typical especially of N Scotland, but only one site for each habitat.
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: Site selection not described. Likely to be opportunistic e.g. on experimental farms. Only one site per habitat. Plot areas probably chosen subjectively

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: One treatment per site, no replication. There was a degree of randomisation in combinations of livestock individuals and grazing period across these and the sites reported in Grant (1985).
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Stock type, numbers and grazing periods given
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: Dietary preference study rather than effects on composition *per se*. Grazing periods only six days, which may be long enough to assess preferences, but perhaps not as proportions vary in response to longer term grazing. There is a possible confounding factor in that the blanket bog was ungrazed for 18 months prior to a spring grazing period, due to higher than expected utilization rates in first two periods.
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: As stated above the blankeg bog plot had an extended period of no grazing which would have allowed some species to recover more than in a more frequently grazed situation
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Likely to be broadly representative, but only one site for each habitat, both in N Scotland
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐-	Comments: In terms of livestock type, although cattle less likely to be grazing on blanket bog, at least in England. The length of grazing period are artificial, for experimental purposes. The resting of the blanket bog

		for a year is not representative of typical grazing units which tend to be grazed annually.

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: Diet from fistulated animals, so direct measurement of what is consumed. Biomass for cutting of random quadrats. Botanical sampling point quadrats, from groups of 50 contacts using a restricted random procedure (transects). IN analysis some grouping of material necessary especially in dietary measurements. Dead material treated together and not separated into species of groups.
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: Grazing periods all similar.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐NA	Comments: Since it is a dietary study longer term effects on vegetation were of less relevance.

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	☐NA	Comments: No comparison groups as such – one treatment plot

in important confounders at baseline?		
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
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: Analysis of variance on transformed data used to analyse differences in diet components between periods and to measure variation between and within animals of the same species. Anova and Cochrane's t-test used to compare sheep and cattle diet, where within and between animal variation was similar (anova) or different (t-test). Kulczynski's similarity co-efficient used to compare sheep with cattle diets and sheep and cattle diets with sward composition.
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 t-tests
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: No site replication, but potential confounding of animal and site effects reduced through randomization of animal and period combinations
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: Sites likely to be fairly typical but only one site per habitat.

Grant, S.A., Suckling, D.E., Smith, H. K., Torvell, L., Forbes, T. D. A. & Hodgson, J. (1985). Comparative studies of diet selection by sheep and cattle: the hill grasslands. Journal of Ecology 73, 987-1004.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: _____Uplands_____

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

Review Question	
Study Citation	Grant, S.A., Torvell, L, Common, T.G., Sim, E.M. & Small, J.L. (1996a). Controlled grazing studies on *Molinia* grassland: effects of different seasonal patterns and levels of defoliation on *Molinia* growth and responses of swards to controlled grazing by cattle. Journal of Applied Ecology, 33, 1267-1280
Study Design Category	2
Assessed by & when	D Martin 13/11/12

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: *Molinia* dominated moorland. Typical soils and management history are given
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: Sites are *Molinia* dominated on peaty or surface water gleys. Low –mid altitude hill land.
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: Cutting treatment imposed at the tussock scale and replicated, so likely to be representative. The grazing areas represent examples of the habitat over different geologies, but with similar soil types, although peat depth varies.

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: Tussocks selected at random for cutting treatment, from within a selected area – selection not outlined. Treatments imposed at tussock scale so limits variability between sampling areas. Cattle Grazing plots probably selected subjectively at each site. Main treatments un-replicated, although small comparison grazing exclosures were replicated at each site.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Cutting and grazing treatments well described. Only two levels of defoliation – no clear basis given for the levels, but intuitively make sense as relatively light and heavy grazing levels.
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: Cutting imposed for 3 years and grazing for 6. This is likely to be adequate, particularly for grazing effects (on plant nutrient and carbohydrate status). Potentially more variability on the grazing treatment as animals adjusted to achieve target sward heights.
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: No
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: Clipping treatment is experimental, designed to simulate grazing, but is an artificial treatment. Summer cattle grazing is a typical

		management regime, but the actual levels will vary, and spatial variability likely to be different on grazing units, which will be larger than the experimental plots.

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, with tillers chosen at random for leaf measurements, or stratified random by quarter or fifth of plot. Veg point quadrat measurements again at random spacing on restricted random transects. Sample sizes at one site (Bell Hill) larger, reflecting larger plot sizes?
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Measures of Plant nutrient and water-soluble carbohydrate status. In the grazing experiment leaf lamina length and extension were measured on 30-50 random tillers, accumulated leaf growth in the season following last grazing period, biomass of ungrazed, one season ungrazed and open area, including live and dead fractions and different species. Floristic composition from point quadrats.
3.4 Were outcomes relevant? If surrogate outcome variables/measurements were used, did they provide a reliable indication of the scale and direction of the important effect(s)?	☐++	Comments: Yes
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes – cutting treatments in place for three years, grazing treatments for 6. Grazing exclusion (control) in place for same period.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐+	Comments: Six years is reasonable time for grazing experiment to influence growth characteristics and vegetation composition.

Section 4: Analyses		
4.1 Were exposure and comparison groups		Comments: Similar vegetation types. Some

similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	differences between sites in altitude, underlying geology and grazing history (ungrazed vs lightly cattle grazed) at the grazing experiment sites. Within sites, the grazed treatment area and ungrazed fenced areas likely to have been similar at start. Baseline measurements of lamina length made at Cleish (but not Bell Hill) before grazing treatments commenced.
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?	☐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: Analysis of Variance, trends over time in plot means using within-plot errors to assess treatment responses. Principal Component Analysis of floristic composition. In clipping treatment sites were analysed separately due to different tussock sizes at start. There is within-site replication.
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 and standard error of means generally given. Cutting experiment means based on 4 observations. Variation in tussock size at start results in large errors. Possibility of type 1 error in significance testing?
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 described experiment, but some differences between plots at start, and variation in tussock sizes including between sites leading to large errors in tussock means.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?	☐+	Comments: Sites reasonably representative, would benefit from more sites given the tussock variation recorded.

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

Evidence Table

Name of Evidence Review:	Uplands
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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

Study details	Authors	Grant, S.A., Torvell, L, Common, T.G., Sim, E.M. & Small, J.L.
	Year	1996
	Aim of study	To identify seasonal patterns and levels of defoliation of *Molinia* that are compatible with sustainable plant production, and effects of 6 years of controlled grazing on herbage production of *Molinia* and other grasses on floristic composition.
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	*Molinia* dominated moorland. Typical soils and management history are given
	Eligible population	Sites are *Molinia* dominated on peaty or surface water gleys. Low –mid altitude hill land.
	Inclusion and exclusion criteria	Molinia dominant, well developed tussock (recently ungrazed or lightly grazed).

	Setting	Cleish Hills, Fife, Sourhope Research Station, Borders and Bell Hill, Borders, all Scotland. All in altitude range 230-450m.
Methods of allocation to intervention/control	Methods of allocation	Tussock clipping treatments assigned at random within blocks of tussocks. Blocks probably chosen subjectively to be representative of site. Grazing applied to plots again chosen to be representative, but not randomised. May have used pre-existing plots.
	Intervention description	Clipping treatments on previously ungrazed tussocks, with two levels of defoliation and four timing treatments. Six replicates of tussock blocks at two sites. Grazing treatments at plot scale with two target defoliation levels (same as cutting) from summer cattle, at two sites.
	Control/comparison description	Clipping compares two levels and timings – no control as such. Grazing experiment has ungrazed exclosures in each plot.
	Sample sizes	Clipping – six tussocks per site, dropping to four as two removed at end of first season for chemical analysis. Post clipping tussock measurements from 4 tussocks. Grazing – lamina lengths on 100 leaves per plot on 30-50 random tillers. Control – 4-5 cages per site. Biomass from 8-10 quadrats. Veg composition from 16 point quadrat locations.
	Baseline comparisons	In grazing expt mean lamina lengths measured in season before grazing commenced. Leaf extension rates prior to grazing measured at one site. Floristic composition measured at start.
	Study sufficiently powered	No power analysis presented. Errors of tussock measurements in clipping experiment large and means based on 4 observations so may be unreliable.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	In clipping, tussock characteristics (above ground biomass, lamina mass, tussock number and weight). In the grazing experiment leaf lamina length and extension were measured on 30-50 random tillers, accumulated leaf growth in the season following last grazing period, biomass of ungrazed, one season ungrazed and open area, including

significance)		live and dead fractions and different species.
	Secondary outcome measures	Measures of Plant nutrient and water-soluble carbohydrate status. Floristic composition from point quadrats
	Follow-up periods	Cutting treatments in place for three years, grazing treatments for 6. Grazing exclusion (control) in place for same period.
	Methods of analysis	Analysis of Variance, trends over time in plot means using within-plot errors to assess treatment responses. Principal Component Analysis of floristic composition. In clipping treatment sites were analysed separately due to different tussock sizes at start. There is within-site replication.
Results		In the clipping experiment the regular defoliation (June, July, August) had the greatest effect on subsequent tussock growth, on weight and tiller number (an expected adaptation to increased grazing would be to produce more tillers), particularly at the higher biomass removal treatment. Single annual cuts only reduced biomass at the higher rate and when done later in the season. There are quite large between site differences in tussock characteristics, both physical and chemical. Where starch content was lower at Cleish it was reduced further at higher defoliation levels. Total water soluble carbohydrates were reduced at both sites at the higher clipping level. In the grazing experiment, rates of leaf extension were generally greater at 33% utilisation. In the post experiment measurements (in the following season) accumulated leaf growth was affected by both previous grazing treatments, with the higher utilisation rate having the greatest effect. In the biomass results grazing changed the balance of Molinia to other grasses, with Molina having more biomass than other grasses in ungrazed plts, other grasses having more biomass at the higher utilisation level, and similar biomass at the lower utilisation level. Grazing had a significant effect on basal internode size. There was also a site effect with different growth potential between sites. Starch and carbohydrates were higher in grazed plots compared with ungrazed, but different tiller sized meant that amounts per tiller were greatest in

		ungrazed. PCA indicates different trends in floristic composition. Despite being initially similar, species number increased in grazed plots, with no or negative change in ungrazed plots. Molinia cover appeared to be levelling off at 55-60% ground cover at 33% utilisation levels, whilst there was a continuing downward trend at 66%. Mechanisms are reduced competition for light, and re-distribution of nutrients by making them available to other species through dung and urine
Notes	Limitations identified by author	Variability of tussock size leading to large errors associated with tussock measurements.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Further work on mechanisms for lack of tillering on Molinia, and role of excreta in creating spatial heterogeneity. Effect of winter grazing on other grasses that have benefitted from summer Molinia defoliation. Evaluation of conservation value to animal populations.
	Sources of funding	Scottish Office Agricultural and Fisheries Dept.

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	Grant, S. A., Torvell, L., Sim, E. M., Small, J. L. & Armstrong, R. H
	Year	1996
	Aim of study	To investigate the prevention of increases in *Nardus* in grasslands through the controlled grazing management of domestic herbivores.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	*Nardus* communities not described in detail, but some general trends and historic management and research presented.
	Eligible population	The study area was previously used in a grazing experiment (Grant et al, 1985)
	Inclusion and exclusion criteria	Plot selected subjectively, used for previous study. Plot size dictated by need to maintain a minimum number of animals

	Setting	Cleish Hills, Fife, Scotland. 280-290m
Methods of allocation to intervention/control	Methods of allocation	Subjective – Need to maintain a minimum of three animals and achieve target sward height meant the only one cattle and two sheep plots were feasible
	Intervention description	Stocked throughout the growing season to achieve an inter-tussock sward height of 4-5cm cattle and 3-4 and 4-5cm sheep. Sward measured twice-weekly. A second experiment had three goat treatments (4-5cm, 5-6cm, 6-7cm) and a sheep control (4-5cm).
	Control/comparison description	No comparison group as such, though sheep 4-5cm seen as control in goat expt.
	Sample sizes	Sheep plots 0.3ha and cattle plot 2.3ha. Goat experiment 0.15ha plots. In each plot multiple measurements of variable are made, e.g. 10 biomass quadrats three times per year, utilisation on 100 tillers, forty measurements of lamina grazing etc.
	Baseline comparisons	Cattle and sheep plots were shown to have similar *Nardus, D flexuosa* and broad-leaved grasses cover, but slightly different in sedges and forbs. They are reasonably close in ordination space for 1984 data. Plots had been burned to remove dead material before the experiment.
	Study sufficiently powered	No replication, low power to detect significant change.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Biomass of different plant groups, Nardus tussocks and uprooted vegetation. Nardus utilisation and tiller growth, and nutrient reserves. Floristic composition and change.
	Secondary outcome measures	

	Follow-up periods	Treatments in place for 5 years
	Methods of analysis	Lack of replication meant that plot and treatment errors confounded, so analysis largely trends in plot means over time, based on t-tests using SEDs calculated for pooled variance for all plots. PCA used to test change in floristic diversity over time.
Results		Total biomass and live mass were greater on the cattle plot than either sheep plot. *Nardus* biomass was greater on both sheep plots than the cattle plot at the end of the experiment, and had increased in both sheep treatments. The weight of uprooted *Nardus* (measured in one year) was greatest in the cattle treatment with no significant difference between sheep plots, and the weight of uprooted fine-leaved grasses was greatest in the heaviest sheep grazed plot. Grazing on *Nardus* tillers was highest in the cattle plot, and lowest in the light sheep grazed plot, with utilisation falling over time on the sheep grazed plots. Growth was greatest in the light sheep grazed plot than the other two, and there was a significant increase in the growth rate over time. Although there were differences between years reflecting growing conditions, tiller base weights and total water soluble carbon (TWSC) were consistently lower in cattle grazed plots. Whilst *Nardus* cover was initially similar at 55%, it had declined to 30% on cattle and 86% and 72% on the sheep 4.5cm and 3.5cm treatments respectively. Broad-leaved grasses increased in the cattle and light sheep plot. *D flexuosa* declined on the heavy sheep-grazed plot. Similarly with goat grazing length of grazed Nardus leaf was positively related to height of between tussock grasses (grazing severity). Growth rates were inversely related to grazing severity. Cattle grazed *Nardus* more readily than sheep. Sheep were more likely to graze *Nardus* when preferred grasses were shorter. Sheep grazed less *Nardus* over time as dead material accumulated. Rate of leaf extension of *Nardus* is about half that of *Agrostis* species, suggesting *Nardus* is not prominent due to competitive vigour but through avoidance. In cattle grazed plots *Nardus* decreased in cover and other grasses either

		increased or stayed the same. Reduction in tiller base weight and selective uprooting by cattle are likely to have played a part. Levels of utilisation which lead to reduced *Nardus* cover and increased *Agrostis* and *Festuca* can be achieved by cattle and goats.
Notes	Limitations identified by author	
	Limitations identified by review team	No replication
	Evidence gaps and/pr recommendations for further research	Exploration of Nardus utilization in mixed sheep and cattle grazing regimes
	Sources of funding	Scottish Office Agriculture and Fisheries Dept.

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	Grant, S. A., Torvell, L., Sim, E. M., Small, J. L. & Armstrong, R. H. (1996b) Controlled grazing studies on Nardus grassland: effects of between tussock sward height and species of grazer on Nardus utilisation and floristic composition in two fields in Scotland. Journal of Applied Ecology 33, 1053-1064
Study Design Category	2
Assessed by & when	D Martin 11/12/12

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* communities not described in detail, but some general trends and historic management and research presented.
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 study area was previously used in a grazing experiment (Grant et al, 1985)
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: Plot selected subjectively, used for previous study. Plot size dictated by need to maintain a minimum number of animals

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: Subjective – Need to maintain a minimum of three animals and achieve target sward height meant the only one cattle and two sheep plots were feasible
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Stocked throughout the growing season to achieve an inter-tussock sward height of 4-5cm cattle and 3-4 and 4-5cm sheep. Sward measured twice-weekly. A second experiment had three goat treatments (4-5cm, 5-6cm, 6-7cm) and a sheep control (4-5cm).
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: 5 year experiment. Allows for grazing effects to be identified against fluctuations in productivity due to weather
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: Not reported, unlikely.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: The description of the vegetation including between-tussock sward is typical of the habitat, however only one small site used.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The habitat is typically grazed, an grazing more likely to take place in summer, particularly with cattle.

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: Biomass of between tussock grasses from 10 random 20x40cm quadrats three times per year, and sub-sampled. Biomass and composition of Nardus tussocks from removal of five random tussock per plot at end of grazing each year. Weight of uprooted shoots of different spp from 10 random 1m quadrats in two years. Nardus utilisation estimated from random tillers (40-100, standardised at the latter after first two years, five tillers at 20 restricted random locations). Forty measurements of grazing severity (lamina length). Leaf extension growth by protecting 30 tillers from grazing. Measurement period varied in duration. Plant chemistry also sampled from three sampling periods per year. Floristics sampled from a min of 25 point contacts at each of 16 locations to obtain percent cover. Essentially same measurements in both experiments, but sample sizes differed
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, although there was some variation between years and between experiments
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: Mainly direct measurements of grazing impact and floristics.
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: Treatments in place for 5 years – long enough to establish grazing effects.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]?	☐++	Comments: No comparison group as such, though sheep 4-5cm seen as control in goat expt. Cattle and sheep plots were shown to have similar *Nardus, D*

Were there any differences between groups in important confounders at baseline?		*flexuosa* and broad-leaved grasses cover, but slightly different in sedges and forbs. They are reasonably close in ordination space for 1984 data.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: No replication
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: Lack of replication meant that plot and treatment errors confounded, so analysis largely trends in plot means over time, based on t-tests using SEDs calculated for pooled variance for all plots. PCA used to test change in floristic diversity over time.
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: Significance of most differences given at $p<0.05$
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: Non replicated
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: Although small scale and unreplicated, measurements were robust and the site is likely to be typical in floristic composition to much acid grassland, although climatic and growth conditions will vary.

GRANT, S. A., SUCKLING, D. E., SMITH, H. K., TORVELL, L., FORBES, T. D. A. & HODGSON, J. 1985. Comparative studies of diet selection by sheep and cattle: the hill grasslands. *Journal of Ecology,* 73**,** 987-1004

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	Grant, S.A.
	Year	1968
	Aim of study	To study the regeneration of heather under a variety of conditions on areas burned as part of a management programme, with and without grazing.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Not described, other than indicating it is heather moorland with burning as part of the management.
	Eligible population	Study areas described in terms of altitude, aspect, slope, soil and heather age. No general vegetation information.

	Inclusion and exclusion criteria	Areas all have high heather cover (at least 75% at start) and are managed by burning as part of the normal management.
	Setting	Scottish Border uplands, and Perthshire and Angus in eastern Scotland. All sites between approx 200m and 600m altitude with all aspects represented.
Methods of allocation to intervention/control	Methods of allocation	Survey of sites all receiving the same treatment – burning. Not clear how the sites were selected, other than they were due to be burned as part of management cycle. The paper suggests all burned in spring, as timing can have important effect on regeneration. There may however be considerable variation in actual timing and, more importantly, conditions at burning.
	Intervention description	Controlled spring burning, with prevailing grazing regime.
	Control/comparison description	Largely a Survey rather than controlled experiment. Each site has an exclosure with no livestock grazing.
	Sample sizes	30 sites with one grazed and one ungrazed sample area.
	Baseline comparisons	Pre-burning heather cover, and post-burn conditions.
	Study sufficiently powered	No power analysis. Study is largely observational rather than analytical.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Heather heights and species composition including heather cover.
	Secondary outcome measures	Rate of heather change over time, observations of grazing effects.
	Follow-up periods	5-8 years

	Methods of analysis	No statistical analysis, just presentation of the observations, including the time taken to achieve 50% heather cover.
Results		A range of factors (firing process, plant characteristics, site characteristics, grazing patterns and climatic factors) influence heather regeneration. One or more factor may have over-riding importance in any one year. There was a tendancy for time taken to reach 50% cover to be longer for older heather at burning. Regeneration of young heather was always quicker as more takes place from shoot than in older heather. Grazing by hare's and grouse had a significant effect at some sites- reducing the difference between the open and enclosed areas. Sheep grazing on burned areas tended to be higher in early years when heather was short, falling over time. This effect varied, depending on factors such as wetness, proximity of better grazing on grass, or surrounding tall heather restricting movement. Whilst most sited achieved 50% heather cover by year 5, but about 25% of sites had not achieved this level by end of the study. Many site factors such as slope, soil and moisture influence regeneration. Trampling emphasised cotton-grass humockiness compared with exclosures, and treading caused the break-up of uncolonised peat surface. Treading can however also help to consolidate soil surfaces, with seedlings thickest along sheep-trods at some sites.
Notes	Limitations identified by author	
	Limitations identified by review team	Largely observational, little analysis. Limited exploration of grazed/ ungrazed difference.
	Evidence gaps and/pr recommendations for	

	further research	
	Sources of funding	

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? 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	Grant, S.A. (1968) Heather regeneration following burning: a survey. Grass and Forage Science, 23, 26-32
Study Design Category	2
Assessed by & when	D Martin 07/11/12

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: Not described, other than indicating it is heather moorland with burning as part of the management.
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: All areas managed by burning. Study areas described in terms of altitude, aspect, slope, soil and heather age. No general vegetation information
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: Not clearly stated, but areas will have been chosen as broadly typical and within the normal burning management of the area.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: Survey of sites all receiving the same treatment – burning. Not clear how the sites were selected, other than they were due to be burned as part of management cycle. The paper suggests all burned in spring, as timing can have important effect on regeneration. There may however be considerable variation in actual timing and, more importantly, conditions at burning.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes – objective botanical measures and heather height
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: There is no non-burned comparison. Each burned area has a un-replicated un-grazed exclosure. The exclosures not fenced against grazing by small mammals, grouse etc.
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 weather, soil moisture etc at burning, and variation in burning practice. Countered by the large number of sites. Paper notes that particularly northern sites were often grazed by hares and grouse, resulting in little difference between the open burn and enclosed area.
2.5 Is the setting applicable to the UK?	☐+	Comments: All sites in East Scotland and Scottish Borders. Typical soils and altitude of UK Uplands, but may not reflect western sites so well.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: objective botanical measurements, and time taken to reach 50% cover (random point quadrat). Not sure of quadrat size/ number.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions	☐++	Comments: Yes

likely to have been identified?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	Comments: In the context of the study the main outcomes were assessed – heather regeneration. Objective estimates of grazing pressure in the vicinity may have been helpful
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?	☐+	Comments: Sites burned over a 4 year period, with observation period ranging from 5-8 years, so some variation in length and timing of observation period.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: The upper end of the observation period range is probably long-enough to identify main effects an trajectory.

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 analysis. Study is largely observational rather than analytical.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: No
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: No statistical analysis, just presentation of the observations, including the time taken to achieve 50% heather cover.
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: No

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: Largely observational study, effects of environmental conditions and timing of burning not assessed, and some confounding from non-livestock grazing in the exclosures.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Yes – particularly relevant to eastern moors managed for grouse. Less relevant to wetter western moors which may not be burned or under less systematic burning.

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	Grazing
Review Question	a. Effect of grazing on biodiversity d. timescales for grazing related change

Study details	Authors	Hartley and Mitchell
	Year	2005
	Aim of study	To quantify the interacting impacts of grazing and soil nutrient addition on rates of vegetation change on moorland systems
	Study design	Quantitative experimental
	Quality score	+
	External validity	+
Population and setting	Source population	2 moors in NE Scotland - Grid ref, rainfall levels, soil types, NVC communities (H12), grazing pressure reported
	Eligible population	2 study sites on each moor - 4 blocks selected per site & 4 5x3m experimental plots selected in each block Selection method/ rationale not reported. Plots assumed to be representative of source population, but not reported

	Inclusion and exclusion criteria	N/A
	Setting	NE Scotland
Methods of allocation to intervention/control	Methods of allocation	method of allocation not reported
	Intervention description	4 sites with 4 blocks of vegetation – 2 blocks fenced in 1993 and grazing treatments allocated. Four 5x3m plots in each block to which a range of N, P, K applications were added
	Control/comparison description	Comparison between fertiliser and grazing treatments as described above
	Sample sizes	2 soil cores from each of 64 plots, 3 1x1m quadrats per fertiliser treatment for vegetation data and 2 sward heights per quadrat
	Baseline comparisons	Not reported
	Study sufficiently powered	Power calculation not reported.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Calluna/graminoid cover, calluna height
	Secondary outcome measures	Species composition, Soil variables
	Follow-up periods	Six year treatment – 1993-1999
	Methods of analysis	GLM used to analyse changes in cover and canopy height of calluna and mean annual browsing damage. GLM used to analyse changes in soil properties Vegetation cover data analysed using constrained linear ordination technique

		redundancy analysis (species occurring infrequently removed from analysis to prevent distortion, and effect of site accounted for) GLM used to analyse the effect of site, fencing and fertiliser on plot scores for Ellenberg, suited species and CSR scores
Results		Protection from grazing had a significant impact on calluna cover ($P<0.0001$), as did site ($p<0.05$). In fenced plots, calluna cover increased on all sites by up to 20%. On plots exposed to grazing, calluna decreased by 20-30% on all sites. Fencing had a significant effect on grass cover (decrease) ($p<0.0001$). Grasses on grazed plots showed an average of 20-30% increase in cover at all sites after 6 years. Nitrogen addition decreased calluna cover on grazed plots, but increased its cover on plots protected from grazing. Calluna height increased in fenced plots, and performed better in unfertilised plots after the 6 years By 1999, species composition was beginning to diverge depending on treatment. Calluna, e. nigrum and e. tetralix were more common on ungrazed plots. V . myrtillus more common with no fertiliser and no grazing. R. Squarrosus more common on grazed plots, J. Squarrosus and T. Cespitosum more common on grazed plots, and E. angustifolium more common on ungrazed plots. Graminoids more common on grazed plots
Notes	Limitations identified by author	Large spatial and temporal variability in responses
	Limitations identified by review team	Small scale and geographically limited experiment
	Evidence gaps and/pr recommendations for	

	further research	
	Sources of funding	James Weir Foundation and The Royal Society Edinburgh

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 d. timescales for grazing related change
Study Citation	Hartley and Mitchell (2005)
Study Design Category	Quantitative experimental
Assessed by & when	SUSANNA PHILLIPS 02/11/2012

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: 2 moors in NE Scotland Grid ref, rainfall levels, soil types, NVC communities (H12), grazing pressure
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: Each moor 2 study sites 4 blocks selected per site 4 5x3m experimental plots selected in each block Selection method/ rationale not reported. Plots assumed to be representative of source population, but not reported
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: Vegetation variables measured at 3 randomly selected sub-plots – method of random selection not reported Estimates of cover taken from point quadrats 2 soil cores per plot taken – method of selection not reported

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 4 sites with 4 blocks of vegetation – 2 blocks fenced in 1993 and grazing treatments allocated – method of allocation not reported
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: Detailed description of experimental design
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: Grazing levels across site as a whole were known, but not on individual 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: Not reported, assumed management intervention as described in experimental design
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: No other interventions reported

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK resource?	☒++ ☐+ ☐- ☐NR ☐NA	Comments: Representative of dwarf shrub dominated habitats (H12) in UK
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ ☐+ ☒- ☐NR ☐NA	Comments: Manipulation of nutrient levels through application of N, P and K

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: Soil organic matter content estimated by loss on ignition 80 pin point quadrat used to estimate species cover Pins used to assess grazed/ungrazed calluna to estimate annual grazing levels Method of calluna height measurements not reported Species composition for whole community recorded visually to nearest 5% and agreed by 2 observers
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 outcomes reported on
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☒++ ☐+ ☐- ☐NR ☐NA	Comments: Appropriate to meet objectives of study

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 used
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Data were recorded on the four sites in May 1993 - 1999
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Six year interval likely to be sufficient to observe a number of changes in variables

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: Not reported
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: 2 soil cores from each of 64 plots, 3 1x1m quadrats per fertiliser treatment for vegetation data and 2 sward heights per quadrat (values from quadrats combined to give mean values per plot)

4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Not reported
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: GLM used to analyse changes in cover and canopy height of calluna and mean annual browsing damage. GLM used to analyse changes in soil properties Vegetation cover data analysed using constrained linear ordination technique redundancy analysis (species occurring infrequently removed from analysis to prevent distortion, and effect of site accounted for) GLM used to analyse the effect of site, fencing and fertiliser on plot scores for Ellenberg, suited species and CSR scores
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 given, means and 1 SE shown graphically for changes in cover
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: Use of subjective measures minimised and observer bias validated. However, method of allocation of treatment not reported
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: Large spatial variations in responses

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? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Hartley, S.E.
	Year	1997
	Aim of study	To investigate whether there are interacting effects of grazing and nutrient inputs on the competitive balance between heather and grasses, and whether grazing and hence its effect is more likely to be concentrated in areas of high plant and soil nutrients.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Source population is Scottish upland moorland mosaic vegetation. Not described in detail but some general comments on trends of heather loss from Scottish moorland.
	Eligible population	The sampling areas are likely to be fairly typical of moorlands in east of Scotland, but again no detailed vegetation description. Background grazing levels are given – moderate – v high. Heather cover declining at both at different rates.

	Inclusion and exclusion criteria	Sites chosen to have a mosaic of heather and grass, and grazed by primarily sheep.
	Setting	Glen Clunie and Glenshee in the Grampian Mountains, Aberdeenshire, Scotland. Altitude 450m to 550m ASL.
Methods of allocation to intervention/control	Methods of allocation	Not clear how selection bias was minimised. Likely that sites were chosen subjectively, but believed to be representative of Soil conditions may be a confounding factor.
	Intervention description	Two blocks at each site were open to grazing and two fenced to exclude grazing mammals. Grazing is simple presence/ absence so subject to external influences. Four nutrient treatments applied at one ungrazed/ grazed combination and four at the other.
	Control/comparison description	Control plots are open to grazing, and unfertilised
	Sample sizes	Two fenced and two unfenced plots at each of four sites. Eight nutrient treatments applied to both a fenced and unfenced plot at each site – so each plot has four treatments applied and each grazing/ nutrient combination is applied once at each site. Vegetation measurements from three $1m^2$ quadrats at each treatment plot – so 192 in total (16 treatment combinations x 4 sites x 3 quadrats).
	Baseline comparisons	Likely to have been chosen as superficially similar, and co-located within the four experimental blocks. However detailed vegetation and soil and plant nutrient comparisons not presented.
	Study sufficiently powered	No power analysis given. Each treatment combination (grazed/ ungrazed x nutrient) has effectively 4 replicates, one from each site, but may be confounded by environmental and background grazing factors.
Outcomes and methods of analysis (inc effect	Primary outcome measures	Calluna height and cover, heather utilisation (proportion of shoots browsed), Calluna canopy.

size, CIs for each outcome and significance)		
	Secondary outcome measures	Nardus plant growth and survival, soil and plant nutrient composition.
	Follow-up periods	Experiment ran for 3 years.
	Methods of analysis	Methods not really described. Statistical tests were used where possible, seem to be simple t-tests or ANOVA. Interactions could be more fully explored using other techniques. Site effects not fully accounted for, although some interactions mentioned as significant.
Results		Heather cover increased more in fenced plots at all sites. It decreased in cover in unfenced plots at two sites, and canopy occupancy decreased in all unfenced plots. Addition of fertiliser at rates of 75 kg/ha/yr N, 12.5 P and 25 K (N level chosen as shown to be 4 times the critical load for heather) increased heather canopy height significantly, but only if protected from grazing. On unfenced sites fertiliser decreased heather cover, but increased it on unfenced sites. Adverse effects of fertiliser addition were therefore only apparent where grazing was present. Nardus showed a marked increase in height on unfenced plots compared to fenced. This would appear to be due to shading effects of heather on fenced areas. It would appear that Nardus is a more effective competitor for nutrients than heather from pot experiments, but this is masked in the field by grazing pressure and soil type. The apparent benefit to Nardus in unfenced areas is an indirect effect of sheep being attracted to fertilised patches and grazing heather. In the absence of grazing, N addition increased heather cover, at the expense of Mat grass. The increase of heather with fencing on all sites and on fertilised plots suggests grazing rather than nutrient addition is a greater influence on vegetation change. High N additions above the critical load no not necessarily lead to heather loss. The results highlight the interaction effect of N deposition and other stress factors.

Notes	Limitations identified by author	The critical load figures derived from work in the Netherlands may not be applicable to Scotland. Any figure may not be widely applicable due to variation in soil type and other factors.
	Limitations identified by review team	Presence/ absence study, limited analysis of site effects.
	Evidence gaps and/pr recommendations for further research	Critical load analysis – examination for Scottish and UK situation?
	Sources of funding	Royal Society of Edinburgh Research Fellowship, NERC.

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? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Hartley, S.E. (1997) The effects of grazing and nutrient inputs on grass-heather competition. Botanical Journal of Scotland 49 (2). 315-324
Study Design Category	2
Assessed by & when	D Martin 2/11/12

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: General comments on the loss of heather from moorland in Scotland, and some background on effects of nutrients from Dutch (lowland) studies. Source population is upland moorland mosaic vegetation. 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: Upland heather moorland – no detailed vegetation description, but the sampling areas are fairly typical of moorlands in east of Scotland. Background grazing levels are given – moderate – v high. Heather cover declining at both at different rates.
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: Two blocks in each of two areas to represent two grazing levels (but neither light). Actual location of study blocks not described – likely to be subjective but chosen to be fairly representative of area. No sources of bias described

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 blocks at each site were open to grazing and two fenced to exclude grazing mammals. Grazing is simple presence/ absence so subject to external influences. Four nutrient treatments applied at one ungrazed/ grazed combination and four at the other. Not clear how selection bias was minimised. Soil conditions may be a confounding factor (**This review concentration on the grazing aspects of the trial – A Pot experiment is not commented on, and less emphasis on solely nutrient addition aspects of field trial**)
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐-	Comments: Grazing is presence/ absence so levels at grazed plots not known. General grazing pressure and utilisation given at the wider moorland level.
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: Three years, so about the minimum required to detect effects for this type of study on vegetation impacts.
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. The grazing treatment is the background levels of the open hill, so subject to external influences.
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: None reported
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Likely to be representative of Scottish moorland (e.g. sites closer to median height than for other parts of UK). On the whole probably reasonable correlation with grazed heather moorland N England, and other parts of UK to lesser extent.

2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Grazed, unfertilised control is likely to be fairly typical of hill grazing practice elsewhere. Experimental nutrient treatments are atypical.

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: Detailed methods not given for many of the measures – soil and plant nutrients, heather biomass etc. Most measures are objective, e.g. vegetation ground cover from 80-pin point quadrats. Detailed canopy measurements made. Scope for some subjectivity in choosing Nardus plants to measure, but three plants chosen per sub-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: Yes
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Probably, in relation to the aims. Effects measured on only Nardus and Calluna, as the co-dominants at the site. Other species, such as mosses, may make in important contribution to ground cover.
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 – direct effects on key dominant and competing species measured.
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 3 years – this will pick up start of trends e.g. in heather cover, but vegetation change will continue over the longer term and there may be step changes not measured in the short-term.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]?	☐+	Comments: Likely to have been chosen as superficially similar, and co-located within the four experimental blocks. However detailed vegetation

Were there any differences between groups in important confounders at baseline?		and soil and plant nutrient comparisons not made.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐-	Comments: No power analysis given. Each treatment combination (grazed/ ungrazed x nutrient) has effectively 4 replicates, one from each site, but may be confounded by environmental and background grazing factors.
4.3 Were the estimates of effect size given or calculable?	☐+	Comments: Mean values with SE of difference generally presented for nutrient analyses, mean values +/- SE for measures such as height, and proportions of ground cover, canopy occupancy etc.
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: Methods not really described. Statistical tests were used where possible, seem to be simple t-tests or ANOVA. Interactions could be more fully explored using other techniques. Site effects not fully accounted for, although some interactions mentioned as significant.
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 analyses of variance where it has been done.
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: Grazing is presence/ absence – not controlled. There is likely to be site effects due to different soils etc not accounted for in the analyses.
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: Will be particularly relevant to East of Scotland. Broadly generalisable but there is variation in climate influences, soils grazing pressures and N lading across the UK uplands.

nationally (i.e. habitat, species)?		

Name of Evidence Review: Uplands Evidence Review

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

Review Question	
Study Citation	Hester and Baillie. 1998. Spatial and Temporal patterns of heather use by sheep and red deer within natural heather/grass mosaics. Journal of Applied Ecology 35 772-784 Hester, Gordon, Baillie and Tappin. 1999. Foraging behaviour of sheep and red deer within natural heather grass mosaics. Journal of Applied Ecology 36 133-146
Study Design Category	Non-randomised controlled trials/controlled before and after studies
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 UK uplands. There is reasonable description of the vegetation type and its context.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: The trial plots were representative of grassland heather mosaics in the uplands. Other Upland habitats were not considered.
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 site was subjectively selected and there is a risk of selection bias.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: There is some selection bias and little indication of how selection bias was minimised.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Yes. Change in vegetation type, sward height, bare ground. Observations in animal behaviour on a timed basis.
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 grazing type. Impact of rabbits occurred. There was however little consideration of the impact of previous parts of the experiment on the resultant animal behaviour.
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. Other significant factors were not discussed. These included the small size and heavy grazing pressure within plots. Also the experiments were very short and seasonally limited. The nutritional needs of the animals in the trial was not considered. Were they hungry? . The sheep had not lambed/were not suckling and their nutritional needs would therefore be very different from a breeding hill ewe. Lack of observation of grazing in the dark means that night time grazing could confound the results. Little consideration give to the disruptive influence of earlier parts of the experiment- grazing animals prefer to take previously grazed vegetation and avoid dead material. In summary: there were too many variables and shortcomings to confidently apply these observations beyond the trial plots and into the wider environment.

2.5 Is the setting applicable to the UK?	☐+	Comments: Study completed in UK Good application to heather-grassland mosaics but not to other 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: Vegetation sampling appeared reliable. However very short observation periods significantly reduces confidence.
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: No- observations were only conducted over short periods . This is an incomplete analysis of the influence of vegetation pattern on grazing.
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: No. These were very short experiments. At least in terms of observation period – experiments ran for 5 yrs

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.	☐-	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. Especially the number of

Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?		animal behavioural observations. The study is poorly powered relating to timescales- these are very short experiments run over a few days. The study is poorly powered when the low number of plots is considered. Overall this is considered as poorly 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 analysis and only a few of the possible explanatory variables were identified.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Yes
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐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: Plot selection was subjective and there were 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: It would be difficult to confidently apply most of the results to any other site in the uplands. A list of the main issues is provided in the evidence table.

Evidence Table

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

Study details	Authors	Hester and Baillie 1998/Hester, Gordon, Baillie and Tappin 1999
	Year	1998 & 1999
	Aim of study	To investigate the influence of vegetation pattern on grazing of heather moorland by red deer and sheep and make observations on grazing behaviour
	Study design	Non-randomised controlled trials.
	Quality score	-ve
	External validity	-ve
Population and setting	Source population	Extensive mosaics of acidic grassland and heathland in the uplands
	Eligible population	Six exclosures erected on a mosaic of Calluna/Festuca-Agrostis grassland
	Inclusion and exclusion criteria	Not described.
	Setting	Mature heather moorland at Glensaugh, Scotland
Methods of allocation	Methods of allocation	Not described.

to intervention/control	Intervention description	Selection of site to typify varying grassland/heathland mosaics followed by introduction of grazing animals (red deer and sheep)
	Control/comparison description	The bulk of this work was an observational experiment measuring animal behaviour but there were two ungrazed plots acting as controls
	Sample sizes	Six exclosures of 1 hectare Observation of animal location and behaviour over short periods during two 4 week grazing periods. Analysis of grazing impact on shoots, dung counting on all plots.
	Baseline comparisons	Baseline comparisons to 2 control plots.
	Study sufficiently powered	Not described but unlikely to be sufficiently powered: a small number of plots with intensive observations over short periods within a short grazing season.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Grazing behaviour and impact of deer & sheep in varying mosaics of acidic grassland and heathland.
	Secondary outcome measures	
	Follow-up periods	None listed .
	Methods of analysis	Statistical analyses conducted- mostly means and standard errors of difference.
Results		Both deer and sheep showed a selective preference for grassland. The plots contained only 15%

		grassland but the grazing time was equally spread over the two habitats. Calluna utilisation is greatest in proximity to grassland. This is seen in deer and sheep although in autumn Calluna utilisation by deer is greater (the experiments did not look at winter utilisation). Even when the actual Calluna utilisation varied, then the proximity to grassland was still influential. More fragmented vegetation therefore showed a greater proportion of Calluna use. Autumn grazing impact on Calluna is heavier as the grazing value of grassland declines. Sheep and deer appeared to consume similar amounts of Calluna during summer, but deer may consume more in autumn and winter and can therefore be more damaging than sheep. Sheep prefer to graze smaller patches of grass whilst deer prefer larger ones, or are less selective. Thus the impact on a grass heath mosaic varies with type of grazing animal. Little Calluna was grazed on the downslope edge of patches- animals did not graze facing down hill. Impact around paths was very noticeable with sheep. Discontinuous grazing will have lesser impact on vegetation as it allows some re-growth of vegetation. Trampling activity and impact due to lying down can be significant in Calluna loss. At low grazing levels trampling impact on Calluna is greater than herbivory. Deer move and lie down more in Calluna whereas sheep prefer paths and lie in grassland. As dunging occurs following periods of inactivity, sheep preferentially dunged on grassland whereas deer preferentially dunged in Calluna. Preferential dunging by sheep on grasslands may impact on their productivity. There was no evidence of sheep and deer interaction which influenced grazing activity. Where herbivore activity was high the work showed that a simple count of number of shoots grazed was an inadequate measure of herbivore activity.

Notes	Limitations identified by author	There was acknowledgement that the growth of heather would be changing as the trial progressed. This would also apply to the grasses. There was acknowledgement of rabbit impacts. If vegetation became short then the animals were removed before the conclusion of the grazing period.
	Limitations identified by review team	The plots were very small and stocked at a very high grazing rate (12 ewes/hectare & 8 red deer/hectare). This might not be representative of grazing in much of the uplands. The experiments were very short with grazing periods of 4 weeks in summer and 4 weeks in Autumn. Grazing behaviour over a whole season would be more meaningful in this context. Animals were limited to grazing small areas within the plots and a limited type of vegetation. Animal behaviour outside plots would be impacted by shelter, weather, moisture in vegetation, palatability of different vegetation in mosaics, disturbance etc. There were issues with applying the breed and age of sheep to sheep farming in the uplands. The sheep had not lambed/were not suckling and their nutritional needs would therefore be very different from a breeding hill ewe. The experiments were so short that it would be possible that the sheep were not hungry- what were they eating before and what were their nutritional needs? There were no observations of animals during hours of darkness when significant grazing can occur. Little consideration give to the disruptive influence of earlier parts of the experiment- grazing animals prefer to take previously grazed vegetation and avoid dead material. The experiment was not modified to remove this bias. In summary: there were too many variables and shortcomings to confidently apply these

		observations beyond the trial plots and into the wider environment.
	Evidence gaps and/pr recommendations for further research	Address the limitations identified.
	Sources of funding	NERC

Evidence Table

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

Study details	Authors	S L Hetherington
	Year	2000
	Aim of study	The use of self-help feed blocks as an aid to grazing and vegetation management of semi-natural rough grazing.
	Study design	1
	Quality score	= QA 5.1 Randomised areas with randomised quadrats and treatments +
	External validity	=QA 5.2 Cambrian mountains have higher rainfall than English Uplands
Population and setting	Source population	4 sites located in the Cambrian mountains ESA. Selected on farms with an ESA agreement imposing stocking units on enclosed land managed as a single unit covered in mosaics of Calluna, Nardus and Vaccinium
	Eligible population	Well representative of the local population but Cambrian mountains have considerably higher rainfall than English uplands. Use of feed blocks as a supplementary feed was normal practice at a locally convenient access point for the farmer
	Inclusion and exclusion criteria	Selected on farms with an ESA agreement imposing stocking units on enclosed land managed as a single unit covered in mosaics of Calluna, Nardus and Vaccinium

	Setting	4 sites located in the Cambrian mountains ESA.
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	2 sites, the normal practice was maintained. 2 sites, feed blocks were strategically placed on areas dominated by Nardus/Vaccinium. Strategic placement conducted at site level, 1 feed block being placed at a number of pre-defined points (feeding stations). Strategic placement applied to sites 2&3 and normal practice to sites 1&4. Strategic treatment applied twice only in April/May & Oct/November 1999.
	Control/comparison description	
	Sample sizes	Baseline data on vegetation composition made on each of the 4 sites, 4 sq m quadrats were placed in Calluna dominated areas and on the interface between Calluna and Nardus communities. On each site, a number of 50sq m experimental areas in the Nardus/Vaccinium communities were identified. On the 2 sites with strategic placement, 6 experimental areas were chosen. Blocks were placed at the centre of 3 and the remaining 3 had no blocks. On the 2 normal practice sites only 3 experimental sites were nominated and remained without feeding blocks..
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Each experimental area had 12 randomly placed 4 sq m quadrats, 3 fenced to prevent grazing. Comparison was made of composition measurements and grazing with 3 of the unfenced quadrats in each area.
	Secondary outcome measures	
	Follow-up periods	

	Methods of analysis	
Results		Calluna was grazed more at interfaces between sward communities compared to quadrats in which it was dominant. Grazing of other key species such as Nardus and Vaccinium was also greater at the interface between communities. The presence of feed blocks can lead to an increase of grazing of the less palatable grasses. The presence of feed blocks led to a general increase in grazing the key species, particularly graminoids but the only *significant* increase observed was for Nardus. The changes in grazing pattern can lead to an increase in deposition of dung near the feed blocks, which could alter soil nutrient availability and could lead to future increased grazing. The results confirm that feed blocks could be used as a passive method of shepherding and also that short term introduction of feeding blocks in undergrazed areas could result in longer term changes in grazing patterns
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Would be good to a follow-up replication
	Sources of funding	MAFF

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 use of self-help feed blocks as an aid to grazing and vegetation management of semi-natural rough grazing. Aspects of Applied Biology 58, 2000. Vegetation management in changing landscapes. S L Hetherington
Study Design Category	1
Assessed by & when	Alison Hiles 7/2/2013

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: 4 sites located in the Cambrian mountains ESA. Selected on farms with an ESA agreement imposing stocking units on enclosed land managed as a single unit covered in mosaics of Calluna, Nardus and Vaccinium
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: Well representative of the local population but Cambrian mountains have considerably higher rainfall than English uplands. Use of feed blocks as a supplementary feed was normal practice at a locally convenient access point for the farmer
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: 2 sites, the normal practice of ad-hoc feed block use was maintained. 2 sites, feed blocks were strategically placed on areas dominated by Nardus/Vaccinium. Strategic placement conducted at site level, 1 feed block being placed at a number of pre-defined points (feeding stations). Strategic placement applied to sites 2&3 and normal practice to sites 1&4. Strategic treatment applied twice only in April/May & Oct/November 1999.

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 sites, the normal practice was maintained. 2 sites, feed blocks were strategically placed on areas dominated by Nardus/Vaccinium. Strategic placement conducted at site level, 1 feed block being placed at a number of pre-defined points (feeding stations). Strategic placement applied to sites 2&3 and normal practice to sites 1&4. Strategic treatment applied twice only in April/May & Oct/November 1999.
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: Apparently well-designed but only one season
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 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: Cambrian mountains have higher rainfall than English uplands
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Controls reflect normal practice

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: Objective measurements. Baseline data on vegetation composition made on each of the 4 sites, 4 sq m quadrats were placed in Calluna dominated areas and on the interface between Calluna and Nardus communities within the experimental area. On each site, a number of 50sq m experimental areas in the Nardus/Vaccinium communities were identified. On the 2 sites with strategic placement, 6 experimental areas were chosen. Blocks were placed at the centre of 3 and the remaining 3 had no blocks. On the 2 normal practice sites only 3 experimental sites were nominated and remained without feeding blocks. Each experimental area had 12 randomly placed 4 sq m quadrats, 3 fenced to prevent grazing. Comparison was made of composition measurements and grazing with 3 of the unfenced quadrats in each area.
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:
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: • Frequency of occurrence of Calluna and Vaccinium in 100 cells of quadrats in communities dominated by Calluna and at the interface between communities • Mean occurrence of key spp – presence in 100 10cmx10cm squares in areas with feed blocks and without feed blocks • Total number of cells grazed and/or contain dung in relation to distance of quadrat from the centre of the area (with or without feed blocks)
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 ☐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: Randomised areas with randomised quadrats and treatments DM Small sample size and few sites
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?	☐++ ✔ ☐+	Comments: Cambrian mountains have higher rainfall than English Uplands

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

Evidence Table

Name of Evidence Review:	Uplands
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? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Hill, M. O., Evans, D. F. & Bell, S. A.
	Year	1992
	Aim of study	To examine the effects of long-term grazing exclusion from a number of sites in North Wales, and to assess how far they enable predictions of future vegetation change under sheep removal
	Study design	2
	Quality score	++
	External validity	+
Population and setting	Source population	Grassland communities dominated by various typically upland species in Snowdonia, N Wales. No detailed description but some discussion of previous studies
	Eligible population	The sampled areas are long-term experimental exclosures set up for previous studies. The range of sites are likely to have been fairly representative of communities when originally set up.
	Inclusion and exclusion	As above. Based on existing long-term plots

	criteria	
	Setting	Sub-montane grassland above 350m in Snowdonia, North Wales. Generally high rainfall and a mix of peaty podzol and brown earth soils
Methods of allocation to intervention/control	Methods of allocation	Replicated exclusion plots: twelve plots 2.8x3.7m. Three treatments with four replicates. Treatment allocation was randomised. However, this paper focuses on long-term exclusion aspects.
	Intervention description	Control was sheep free access. Other two treatments were: sheep excluded in winter; year round exclusion. Free-ranging sheep grazing levels were calculated, ranging from 5 sheep per ha on brown earths, to 1.9 sheep per ha at altitude. Not presented for each site, and will have changed over time. Treatments continued for varying lengths of time with only two following this management by 1975. **However, this paper focuses on the grazing exclusion plots,** although it does follow the treatment effects at the two sites where they persisted. The other sites were ring fenced at this time to create larger ungrazed areas. The two sites in the original regime were similarly ring-fenced in 1882. Exact duration of treatments at some sites unclear.
	Control/comparison description	Original controls were plots open to the prevailing agricultural grazing regime.
	Sample sizes	Nine study sites with four replicates of three original treatments, including control. Only two sites continued in original form beyond 1975 until 1982. Stock excliusion continued at all sites. Cover estimated from 100 randomly placed pins per plot. From 1981 the ring-fenced plots were sampled destructively in small quadrats, nine 20 x 20cm per plot.
	Baseline comparisons	Blocks varied, as deliberately targeted at different vegetation types. No indication of similarity of plots within blocks at start.
	Study sufficiently	No power analysis carried out

	powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation composition (percentage cover) and change over time. From 1981 measured in terms of biomass
	Secondary outcome measures	
	Follow-up periods	Stock exclusion treatments ran for between 13 and 24 years. The seasonal grazing treatments were in place for a variable amount of time.
	Methods of analysis	Presented largely as mean percentages of cover of species at grazing removal and after a number of years of exclusion. Students t-test of differences at the two longer-term treatment comparison sites, ignoring randomized block design and using means for each treatment.
Results		Sites varied in character at the outset, with Nardus, Molinia and Agrostis/ Festuca grasslands. Long-term change depended on starting point. The long-term sites where winter removal continued showed there was virtually no difference over time between this and year-round grazing. There were however more significant differences in species proportion between the grazing exclosure and grazed treatments, generally increasing over time. Molinia had higher cover in the ungrazed plots and Nardus lower. Ericoids initially expanded through growth of existing bushes, then degenerated. Change was initially rapid in the first 8 years, then slowed. Over the nine sites where stock was excluded species which showed the greatest declines were low growing, including some small sedges and heath rush. Palatable grasses and herbs and ericoids showed the greatest increases. Agrostis/ festuca grasslands on brown earths changed less than more 'heathy' grasslands on podzolic soils where D fexuosa, Molinia or ericoids became more prominent at the expense of

		Nardus, Sheep's fescue and heath rush and other low-growing plants. Few new species appeared other than broad buckler fern on grass litter and rowan along fencelines. Almost all change occurred through clonal spread or growth of individuals. Peaks in vole abundance resulted in dead grass and moss, but no bare ground. They are also contributors to variation in biomass, in the absence of sheep.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Information on invasability of habitats and possible mechanisms. Follow up could include experimental re-introduction of grazing and controlled burns.
	Sources of funding	

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? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Hill, M. O., Evans, D. F. & Bell, S. A. (1992) Long-term effects of excluding sheep from hill pastures in North Wales. Journal of Ecology, 80, 1-13
Study Design Category	2
Assessed by & when	D Martin 10/12/12

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: Grassland communities dominated by various typically upland species in Snowdonia, N Wales. No detailed description but some discussion of previous studies
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 sampled areas are long-term experimental exclosures set up for previous studies. The range of sites are likely to have been fairly representative of communities when originally set up.
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 as per the eligible population. Site selection not described in detail. Chosen to represent a range of upland vegetation and soil types, but 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: Replicated exclusion plots: twelve plots 2.8x3.7m. Three treatments with four replicates. Treatment allocation was randomised. However, this paper focuses on long-term exclusion aspects.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Control was sheep free access. Other two treatments were: sheep excluded in winter; year round exclusion. Free-ranging sheep grazing levels were calculated, ranging from 5 sheep per ha on brown earths, to 1.9 sheep per ha at altitude. Not presented for each site, and will have changed over time. Treatments continued for varying lengths of time with only two following this management by 1975. **However, this paper focuses on the grazing exclusion plots,** although it does follow the treatment effects at the two sites where they persisted. The other sites were ring fenced at this time to create larger ungrazed areas. The two sites in the original regime were similarly ring-fenced in 1882. Exact duration of treatments at some sites unclear.
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: Sites established between 1957 and 1968, and active management continued to early 1970s. Two continued until 1982. After these times the experiment changed to exclude grazing from all plots. There was therefore quite a lot of variation in exposure to treatments, and in length of time of grazing exclosure, ranging from 13-24 years. This is enough time to pick up effects, but exposure varies between different sites and therefore vegetation types. The control plots were effectively lost when grazing excluded from blocks.
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 continuity of grazing exclusion treatments, but it would see that winter exclusion plots were in place for different lengths of time and succumbed to year-round grazing access.
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	☐+	Comments: Treatments kept up for different lengths of time at different sites. It's likely that winter exclusion is the treatment that was abandoned and subject to year round grazing. However it seems that the exclusion treatments have had continuity

unplanned burning)? Were groups treated equally?		
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comment: As there are a number of sites they are reasonably representative of typical grazed upland communities.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Control plots reflected background agricultural sheep grazing levels

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 from 100 randomly placed pins per plot. May not be representative or take account of variation in dominance or different communities within plot. From 1981 the ring-fenced plots were sampled destructively in small quadrats, nine 20 x 20cm per plot. Not stated if objective, but placed to avoid previous locations.
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: Vegetation measurements not made in every year, but no rationale given for sampling pattern. The biomass sampling from 1986 was "mostly lost"
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Main aims were to investigate vegetation change, which it largely does, in terms of percent cover of species. No structural measures or attempts to classify the communities present.
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: Control plots (background grazing) were effectively lost at different times.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Grazing exclusion periods are 13-24 years. Long enough to detect effects of grazing removal, but the period of grazing exclusion varies between sites and therefore vegetation types.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐NR	Comments: Blocks varied, as deliberately targeted at different vegetation types. No indication of similarity of plots within blocks at start.
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?	☐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: presented largely as mean percentages of cover of species at grazing removal and after a number of years of exclusion. Students t-test of differences at the two longer-term treatment comparison sites, ignoring randomized block design and using means for each treatment.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	Comments: p values for t-test results.
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: Fairly large-scale and long-term study, across a range of vegetation types. Treatments randomised. However other treatments and control not maintained, and limited analysis
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?	☐++	Comments: The range of vegetation types are typical of grazed upland areas. It is likely that generalisations can be drawn from the range of sites in this study.

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

Evidence Table

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

Study details	Authors	J Hodgson, T.D.A. Forbes, R.H. Armstrong, M.M. Beattie, E.A. Hunter
	Year	1991
	Aim of study	Comparative studies of the ingestive behaviour and herbage intake of sheep and cattle grazing indigenous hill plant communities
	Study design	2
	Quality score	+ = QA 5.1 The use of a single group of mature, non-reproductive animals to graze the 6 sites in random sequence was intended to minimise the risks of confounding differences between communities and seasons of measurement with between-animal differences and the effects of changes in the physiological state of the animals concerned. There is evidence to suggest that animals subjected to a major change in vegetation type may require a period of several weeks or months before exhibiting similar selective behaviour to that of animals with prolonged experience but the potentially confounding effect on these comparisons of between-animal differences in selective behaviour is not clear
	External validity	:= QA 5.2 The different types of sward are sufficiently described and the findings detailed enough for reasonable generalisation of the findings nationally.
Population and setting	Source population	Upland grassland in southern Scotland

	Eligible population	Perennial ryegrass sward, Agrostis/Festuca, Nardus, Molinia, Calluna vulgaris/Eriophorum vaginatum blanket bog and Calluna Moor sites
	Inclusion and exclusion criteria	Correct sward type. Altitude 240-280m except PRG at 150m
	Setting	Cleish Hills, Fife Forest District and Bell Hill, Wauchope Forest, Roxburgh, Glensaugh Research Station
Methods of allocation to intervention/control	Methods of allocation	The same group of animals was used throughout and the results were accumulated over three years from a random sequence of sites across seasons and years
	Intervention description	NA
	Control/comparison description	NA
	Sample sizes	6 sites, each fenced into 2 adjacent plots each approx 3ha
	Baseline comparisons	NA
	Study sufficiently powered	NR
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 sheep consistently maintained a higher level of extrusa digestibility than the cattle....., reflecting the generally greater degree of selectivity in their grazing behaviour..... The absolute differences were relatively small on all grassy communities, but sheep selected diets of substantially higher digestibility than cattle on the shrub communities. Overall, differences between species in rate of biting and grazing time were relatively small and not significant, the marginally higher biting rate of sheep being counterbalanced by the marginally higher grazing times for cattle, so that daily bites were similar..... However, there was a significant species x community interaction in grazing time, values being 20% lower for sheep than for cattle on the *Molinia* community but 15-25% greater on the shrub communities' 'Four swards were too short for the animals to graze deeper than 16.5cm and on the May 1979 *Nardus* sward no grazing depths exceeded this value. Of the remaining seventeen comparisons, the proportion of records exceeding 16.5cm penetration was significantly greater for sheep than cattle on nine occasions, not significantly greater on four occasions, and significantly less on four occasions.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	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	Moorland Gazing and stocking rates
Study Citation	Comparative studies of the ingestive behaviour and herbage intake of sheep and cattle grazing indigenous hill plant communities. J Hodgson, T.D.A. Forbes, R.H. Armstrong, M.M. Beattie, E.A. Hunter
Study Design Category	2
Assessed by & when	Alison Hiles 31/01/2013

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: : Agrostis/Festuca, Nardus and Molinia grassland communities in Southern Scotland plus 2 dwarf shrub communities (Calluna vulgaris-Eriophorum vaginatum blanket bog and Calluna Moor) For full details we are referred to Grant et al (1985 & 1987)
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: We need to refer back to the Grant et al papers for this – well described there but not in this paper

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: As Grant papers
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: All interventions and measurements were described in minute and exact detail though some depended on reference to the Grant papers
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 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: All in southern Scotland but sward types reasonably typical of upland England also.
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:
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:
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: The use of a single group of mature, non-reproductive animals to graze the 6 sites in random sequence was intended to minimise the risks of confounding differences between communities and seasons of measurement with between-animal differences and the effects of changes in the physiological state of the animals concerned.
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 use of a single group of mature, non-reproductive animals to graze the 6 sites in random sequence was intended to minimise the risks of confounding differences between communities and seasons of measurement with between-animal differences and the effects of changes in the physiological state of the animals concerned. There is evidence to suggest that animals subjected to a major change in vegetation type may require a period of several weeks or months before exhibiting similar selective behaviour to that of animals with prolonged experience but the potentially confounding effect on these comparisons of between-animal differences in selective behaviour is not clear
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?	✔ ☐++ ☐+	Comments: The different types of sward are sufficiently described and the findings detailed enough for reasonable generalisation of the findings nationally.

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

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	c) What changes have taken place under recent reductions and seasonal changes in sheep grazing, and what is the significance of these changes?

Study details	Authors	Hope, D, Picozzi, N, Catt, D. C. & Moss, R.
	Year	1996
	Aim of study	To assess what effects there might be on common semi-natural upland vegetation communities, and on the main wild vertebrate herbivores associated with them, when sheep are removed from large tracts of rangeland in the Scottish Highlands
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	The extent of semi-natural upland vegetation communities and associated wild herbivores
	Eligible population	Eligible population are areas of moorland or large semi-natural enclosures where sheep had been removed. Eleven sites with stock reductions and paired grazed controls. The sites are geographically widespread through the North, East and West Highlands. A wide range of sub-montane grassland, heath and bog communities represented
	Inclusion and exclusion	Sites selected to include the most common upland plant communities, and to encompass a variety of management activities. Sites had to be large enough to assess

	criteria	effects on wild herbivores, including red and roe deer.
	Setting	Range of sites across Scottish Highlands between 100m and 550m in altitude
Methods of allocation to intervention/control	Methods of allocation	There was necessarily a degree of selectivity or opportunism involved in identifying sites with sheep removal. Sample stratified by three bioclimatic regions.
	Intervention description	Removal of sheep. At two sites grazing changed to summer only
	Control/comparison description	Grazing at prevailing farming levels. May have been subject to some variation over time.
	Sample sizes	11 pairs of sites. Most variables sampled on between 6 and 18 plots per vegetation type per study area.
	Baseline comparisons	N/A
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Presence of wild herbivores, vegetation composition and structure
	Secondary outcome measures	Presence of voles
	Follow-up periods	Stock reductions for between 1 and25 years
	Methods of analysis	Difference in vegetation attributes between control and reduced area using Mann-Whitney U tests due to lack of normal distribution in data. Difference in height of structural layers between treatments tested using t-test. Anova used to test differences in patch size in vegetation type and grazing regime. Mann-Whitney also used on pellet groups for different grazing species. PCA used to investigate vole run frequency against

		vegetation attributes.
Results		Differences in cover between reduced grazing areas and control were found most frequently in grasses: Agrostis, Festuca and Molinia. Cover of Carex spp and Deschampsia flexuosa tended to be higher in reduced sheep areas, with heather having higher cover in three sites. Vegetation was usually taller in reduced sheep areas. At sites where sheep had been removed for over 5 years differences were most pronounced in dwarf shrub, grass and moss layers. Patches of heather tended to be larger and grassland smaller on reduced-sheep areas in western and some northern sites. There was a tendency for red deer pellet groups to be more frequent on control areas at western sites, but were variable over all control areas. The difference in frequency of vole runs between reduced sheep and control sites was correlated with the difference in sward height at the same sites. Height and tussock frequency explained most of the variation. Reduced sheep grazing was shown to quickly result in taller vegetation, with few apparent changes in floristic composition. Patches of dwarf shrub-dominated vegetation tended to be larger and grassland smaller where sheep had been reduced. Vole activity was shown to increase as grass height increased above 5cm. Grazing by red deer and continued heather burning limited change in many sites.
Notes	Limitations identified by author	Some differences in approaches to sampling – stratifying by veg type in this study – may result in less marked composition change results compared with other studies
	Limitations identified by review team	One –off survey so limited identification of causality. Sheep grazing levels not quatified, and not clear on what size of reductions took place, and whether there has been complete sheep removal in at least some sites
	Evidence gaps and/pr recommendations for	

	further research	
	Sources of funding	SOAFD

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	c) What changes have taken place under recent reductions and seasonal changes in sheep grazing, and what is the significance of these changes?
Study Citation	Hope, D, Picozzi, N, Catt, D. C. & Moss, R. (1996) Effects of reducing sheep grazing in the Scottish Highlands. Journal of Range Management, 49, 301-310
Study Design Category	2
Assessed by & when	D Martin 16/12/12

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 extent of semi-natural upland vegetation communities and associated wild herbivores
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 population are areas of moorland or large semi-natural enclosures where sheep had been removed. Eleven sites with stock reductions and paired grazed controls. The sites are geographically widespread through the North, East and West Highlands. A wide range of sub-montane grassland, heath and bog communities represented
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++	Comments: As per eligible population. Sites selected to include the most common upland plant communities, and to encompass a variety of management activities. Sites had to be large enough to assess effects on wild herbivores, including red and roe deer.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: There was necessarily a degree of selectivity or opportunism involved in identifying sites with sheep removal. Sample stratified by three bioclimatic regions.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Main variable is presence of sheep grazing
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: May well have been some trespass onto reduced grazing areas. However it is a survey type approach rather than experimental, so estimates of actual grazing pressure made.
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: Climatic variation and therefore growing conditions taken into account through number and distribution of sites. Soils sampled and current and past management characterised. Sites were rejected where reduced sheep and control were too dissimilar.
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes but all sites in N Scotland so likely to be wetter with shorter growing seasons.

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 composition and structure and herbivore dung and signs of vole presence from 2m plots from constrained random sampling, on transects, with 6-18 plots per vegetation type. Patch size assessed on two transects.
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 stated objectives

Were all important positive and negative effects assessed?		
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: The surrogate measure of pellet group counts tested for repeatability by calculating intra-class correlation coefficients for reduced-sheep and control areas.
3.5 Were there similar follow up times in exposure and comparison groups?	☐-	Comments: Reductions were in place for varying lengths of time. Grazing pressure in the comparison areas may have fluctuated over the reduced grazing period.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Variable, but up to 25 years, which would allow longer-term effects to be noted.

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?	☐+	Comments: Grazing regime and vegetation type were the main variables. Vole presence was analysed against veg height and structural attributes.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++	Comments: Difference in vegetation attributes between control and reduced area using Mann-Whitney U tests due to lack of normal distribution in data. Difference in height of structural layers between treatments tested using t-test. Anova used to test differences in patch size in vegetation type and grazing regime. Mann-Whitney also used on pellet groups for different grazing species. PCA used to investigate vole run frequency against vegetation attributes.
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 all tests

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 approach so difficult to control for sources of bias, but uses paired sites with control and stratified by geographical area. However the grazing levels of control may have varied over time
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 cover a range of common upland vegetation types, however grazing conditions and wild grazers reflect mainly Scottish situation

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?
Study Citation	The effect of controlled sheep grazing on the dynamics of upland *Agrostis-Festuca* grassland. Hulme, P.D., Pakeman, R.J., Torvell, L., Fisher, J.M. & Gordon, I.J. 1999. J App Ecol 36
Study Design Category	1
Assessed by & when	D Martin 1/10/12

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.	X++ ☐+ ☐- ☐NR ☐NA	Comments: Basic community described in terms of NVC and dominant grasses, With difference in key grass species that respond to changes in grazing described.
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ x+ ☐- ☐NR ☐NA	Comments: Represent two variants of *Agrostis-festuca* acid grassland, one perhaps more 'moorland' in character with *Nardus* and *Molinia*, and the other maybe more typical of extensively managed enclosed hill grassland. Sampling frame dictated by site availability.
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?	☐++ ☐+ ☐- XNR ☐NA	Comments: Selection of location of plots, or similarity to wider area is not reported. There may be sources of bias, but likely to be fairly representative of the wider area.

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?	X++ ☐+ ☐- ☐NR ☐NA	Comments: Within each block, treatments were imposed randomly. Exclusion plots in each block so likely to encompass more variation at start than the two replicates of other treatments.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++ X+ ☐- ☐NR ☐NA	Comments: Three target sward heights described and could in theory be replicated, although treatments reactive to sward measurements and anticipation of growing conditions so not precise. Small plots not so easy to fine tune as larger fields as each sheep contributes relatively large proportion of the grazing, especially on less productive plots. So ease of maintaining target swards may vary between sites. Wethers rather than ewes – different grazing preferences?
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)	☐++ ☐+ X - ☐NR ☐NA	Comments: Treatment imposed for 6-7 years, so reasonable period of time to allow affects on veg dynamics. Treatment commenced later at Kirkton. Some difficulty of maintain sward heights at Kirkton – 4.5 and 6 cm often lower than target. Swards much taller in early 1995 than target in these two treatments. Stock put and take and sward measurements by different people at each site.
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?	☐++ X+ ☐- ☐NR ☐NA	Comments: No direct contamination, but some difficulty of maintaining sward height differentials – due to inherent variability of system.
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?	☐++ X+ ☐- ☐NR ☐NA	Comments: No other intervention apparent, although there will be seasonal climatic variation which will differ between the two sites.

2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	X++ ☐+ ☐- ☐NR ☐NA	Comments: Sites are in Scotland. The vegetation types are widespread, at altitudes similar to the experiment sites. NVC sub-community from Kirkton is reported as occurring in N England, may be less typical of more southerly uplands.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++ X+ ☐- ☐NR ☐NA	Comments: Broadly – extensive sheep grazing, 'typical' treatment around 1- 1.5 sheep per ha on moorland site. Main difference is use of wethers.

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?	X++ ☐+ ☐- ☐NR ☐NA	Comments: measurements objective – sward heights by HFRO sward stick (maybe differences between observers at the two sites). Veg composition and cover sampled systematically by point quadtrats with standardised min number of contacts to account for different heights. Point qudrats give objective measures of cover.
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)?	☐++ X+ ☐- ☐NR ☐NA	Comments: No floristic measurements in 1994.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++ X+ ☐- ☐NR ☐NA	Comments: Herbage mass measurements abandoned doe to observer variability. Vegetation composition is the key outcome.

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)?	X++ ☐+ ☐- ☐NR ☐NA	Comments: Outcomes relevant – impact on species composition and change is key measure
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	X ++ ☐+ ☐- ☐NR ☐NA	Comments: within site treatments imposed for similar time, although two sites stated at different times.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ X+ ☐- ☐NR ☐NA	Comments: 6 or 7 years of treatments – medium-term exposure as upland habitats are fairly stable and change slowly.

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?	☐++ X+ ☐- ☐NR ☐NA	Comments: At Cleish samples appear similar at baseline – "most" had affinity to typical U4a sub-community, but ordination detects some differenced at outset. Greater variation at Kirkton with elements of wet heath vegetation
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?	☐++ ☐+ ☐- X NR ☐NA	Comments: No power calculation exists. Replication is limited – 2 per treatment at each site, other than the exclosure treatment.

4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ X - ☐NR ☐NA	Comments: Only in terms of movement in ordination space, so relative magnitude of treatment effects.
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?	☐++ X + ☐- ☐NR ☐NA	Comments: Mainly ordination techniques that allow each plot to be shown in ordination space. Repeated measures ANOVA on species – low d.f. for treatment? Significance of time as well as treatment given – both sites show some background change in species across treatments.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ X+ ☐- ☐NR ☐NA	Comments: p values given for the treatment, time and interaction effects on main species
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?	☐++ X + ☐-	Comments: Low replication and some variation between plots at start. But plots treated individually in ordination analysis. Difficulty of imposing sward heights consistently, especially at Kirkton.
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)?	☐++ X + ☐-	Comments: There are general principles in terms of target sward height that can be applied to management of *Agrostis-Festuca* grassland, and takes account of some of the variability in terms of the key grassland species that may replace the more palatable ones. However exact response will vary across the resource due to differences in productivity and composition. The study did not include winter grazing that appeared to control Nardus outside of the plots and used wethers (male) rather than more typical (and likely more selective) ewes.

Evidence Table

Name of Evidence Review:	Upland
Name of Review Sub-topic (if any):	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery?
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	**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:	Source	Methods of allocation: Two	Primary outcome	Differences in	Limitations identified

Hulme, P.D., Pakeman, R.J., Torvell, L., Fisher, J.M. & Gordon, I.J. Year: 1999 Aim of study: To investigate the effects of sheep grazing intensity on the dynamics of Agrostis-Festuca grassland Study design: Randomised block for three grazing treatments, with non-grazed control sub-blocks Quality Score +	population: Upland *Agrostis – Festuca* dominated acid grassland vegetation (NVC U4) Eligible Population: Two variants of above community – one more moorland in character and one more productive Inclusion & exclusion criteria: Setting: Two hill farms in Scotland used for studies of extensive hill livestock systems	blocks of three plots established at each site. Three treatments allocated randomly to one plot in each block. Small ungrazed control exclosure established in each plot. Intervention description: Three sheep grazing treatments implemented by maintaining average summer sward heights of 4.5cm (typical), 3cm (heavy) and 6cm (light). Small sub-plot ungrazed in each block Control / comparison description: Small sub-plot (5mx5m) ungrazed in each block Sample sizes: Two replicates of 0.3ha at each site. Veg measures from twenty systematically placed frame quadrats, on 4 or 5 transects. Variable number of pin traverses to achieve min of 25 hits at each location.	measures: Secondary outcome measures: Species frequency and cover from pin quadrats, sward height measurements. Herbage mass abandoned due to observer variability Follow-up periods: Annual measurements summers 1990-1995 (except 1994). Also earlier baseline (1989) at one site. Methods of analysis: Ordination of species data and trends in composition of each plot over time, presented for each site. Relative movement of each treatment in ordination space give. Repeated measures Anova of individual species.	sheep grazing days required to achieve target sward heights at the two sites. Lower levels of grazing required at moorland site. Changes in species composition over the 7 years small, with few spp lost or gained. At moorland site low sward heights allowed *Nardus stricta* to spread. Where this sp absent at the productive site, mosses increased. Lack of grazing allowed grazing-intolerant grasses to increase. Least change associated with 4.5 cm at productive site, and 6cm at moorland site.	by author: No winter grazing in plots, which may explain difference in response of Nardus to outside of plots. Limitations identified by review team: Low replication, difficulty of achieving and maintaining sward heights, esp in low productivity plots. Evidence gaps and/pr recommendations for further research: Effects of year round inc winter grazing in controlling *Nardus*. Sources of funding: Scottish Office Agric, Env and Fisheries Dept.

External validity: +		Baseline comparisons: Floristic composition and cover measurements and NVC assessments for each plot Study sufficiently powered: No power analysis, low replication			

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? b) What methods of stocking rate calculation, or setting grazing regimes, consistently provide regimes that maintain or restore moorland biodiversity, and what are the key parameters that calculations should include?
` Citation	Hulme, P. D., Merrell, B. G., Torvell, L., Fisher, J. M., Small, J. L. & Pakeman, R. J. (2002) Rehabilitation of degraded *Calluna vulgaris* (L) Hull- dominated wet heath by controlled sheep grazing. Biological Conservation 107, 351-363
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: Wet heath system chosen as there were previous studies on dry heath and bog. General description not given
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: Not described in detail. Degraded and suppressed heather present.
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 represent the wider heft, but some variation reported in terms of the patchiness of heather in each block. Block selection likely to be subjective, but positioned so that treatment plots have similar proportions of heather and grass. Vegetation within blogs described as having closest match to M15.

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 blocks of replicated treatments. Not clear if randomised.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Based on set stocking rates. Summer, Winter and year-round low sheep grazing (0.7 sheep ha^{-1}), and year-round moderate (1.4 sheep ha^{-1}). The control was the existing heft regime of 2.1 sheep ha^{-1}. Each treatment plot had ungrazed fenced exclosure.
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 6 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: None reported.
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐++	Comments:
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Likely to be representative of grazed wet heath.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: The control represented typical commercial hill grazing rates. Treatments other than winter only have similarities to conservation maintenance or restoration grazing regimes.

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: Utilisation and sward height measurements at random points. Heather utilisation following established methods. Vegetation from inclined point quadrats at fixed points.
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: Utilisation, shoot measurements – length, diameter, structural component weights. Vegetation composition.
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: In place for 6 years

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: The plots were set up within each block to have a similar proportion of heath to grass, however some difference in the size of patches between blocks were described. Plots were similar in ordination space in 1989 at start.
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: Heather utilisation and sward height analysed by ANOVA, taking account of time effects. Species relative frequency analysed in different ways, comparing coefficients produced by orthogonal linear contrasts (Genstat) within a randomised block Anova. RDA on floristic data . A range of factors including 'treatment.time' was included. Significant of factors estimated by Monte Carlo permutation.
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 all analyses
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 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: Robust design, community likely to be representative of the wider population

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? b) What methods of stocking rate calculation, or setting grazing regimes, consistently provide regimes that maintain or restore moorland biodiversity, and what are the key parameters that calculations should include?

Study details	Authors	Hulme, P. D., Merrell, B. G., Torvell, L., Fisher, J. M., Small, J. L. & Pakeman, R. J.
	Year	2002
	Aim of study	To determine the management needed to restore degraded heather in a wet heath system where heather loss was known to have recently occurred.
	Study design	2
	Quality score	++
	External validity	++
Population and setting	Source population	Wet heath system chosen as there were previous studies on dry heath and bog. General description not given
	Eligible population	Not described in detail. Degraded and suppressed heather present. Likely to represent the wider heft, but some variation reported in terms of the patchiness of heather in each block.

	Inclusion and exclusion criteria	Block selection likely to be subjective, but positioned so that treatment plots have similar proportions of heather and grass. Vegetation within blogs described as having closest match to M15.
	Setting	Redesdale Experimental Farm, Northumberland. Gleys and shallow peats at an altitude of 300m OD
Methods of allocation to intervention/control	Methods of allocation	Two blocks of replicated treatments. Not clear if randomised.
	Intervention description	Based on set stocking rates. Summer, Winter and year-round low sheep grazing (0.7 sheep ha^{-1}), and year-round moderate (1.4 sheep ha^{-1}). Each treatment plot had a fenced exclosure.
	Control/comparison description	The control represented typical commercial hill grazing rates. The control was the existing heft regime of 2.1 sheep ha^{-1}.
	Sample sizes	Heather utilisation measured on 100 random shoots per plot, sward height at 40 points. Shoot measurements form 60 points in each plot. Twenty sets of quadrat frames, each with 100 pin measurements, in each plot.
	Baseline comparisons	The plots were set up within each block to have a similar proportion of heath to grass, however some difference in the size of patches between blocks were described. Plots were similar in ordination space in 1989 at start. Biomass utilisation and sward height were very similar at start in all plots. These levels were maintained throughout on the control.
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect	Primary outcome measures	Heather utilisation following established methods. Vegetation from inclined point quadrats at fixed points.

size, CIs for each outcome and significance)	Secondary outcome measures	Sward height and heather shoot measurements
	Follow-up periods	Treatments in place for 6 years
	Methods of analysis	Heather utilisation and sward height analysed by ANOVA, taking account of time effects. Species relative frequency analysed in different ways, comparing coefficients produced by orthogonal linear contrasts (Genstat) within a randomised block Anova. RDA on floristic data . A range of factors including 'treatment.time' was included. Significant of factors estimated by Monte Carlo permutation.
Results		Heather prior to fencing and in control experienced offtake in excess of 60% of shoots. Throughout the experiment utilisation was significantly less on all fenced than unfenced area (less than 40% of the grazed control). It was lowest in summer low grazing, and highest in the year-round moderate, but not significantly so. In all treatments heather height increased to over 35% by year 6, and post-summer grass height was similar. RDA showed a clear treatment effect on species composition (variance explained by treatment significant p=0.005). Stock exclusion had the greatest effect, and the winter grazing treatment had different effects from the summer and tear-round treatments. In addition to heather, C nigra, D flexuosa, G saxatile and E tetralix benefitted from reduced summer grazing. Molinia and P erecta were particularly associated with winter grazing. High year round grazing was associated with moss and low-growing speies such as C fontanum. The total relative frequency of heather increased in all fenced treatments. Shoot lengths in 1994 were significantly longer in fenced treatments, but did not differe in diameter. Total shoot dry weights were 5 times higher in fenced than unfenced areas (seven times higher for green portion, four times for woody). Utilisation in all grazed treatments was within the range in which heather is thought to be able to maintain its growth, and resulted in similar responses in height and

		frequency increase between treatments. Reduction in utilisation was not proportional to sheep number reductions, with the greatest reduction in summer only grazing when palatable grass growth is at a maximum. Lack of summer grazing allowed Molinia to thrive, even though heather utilisation was low enough not to affect growth of this species. Summer grazing kept Molinia in check, whilst exclosure increased the competition from heather. From the experiment a stocking rate of between 0.7 sheep ha^{-1} and 1.4 sheep ha^{-1} in a year-round grazing regime will increase the vigour of previously heavily grazed heather on wet heath, whilst a rate of 2.1 sheep ha^{-1} results in continued degradation.
Notes	Limitations identified by author	Setting of stocking rate is not simplistic, and is influenced by spatial pattern of vegetation. Management should be reactive and monitoring is required for accurate management decisions.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	Not stated

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: ___Upland_____

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

Review Question	Compare the behaviour of individual sheep and that of the flock as a whole.
Study Citation	Hunter, R.F. and Milner, C. (1963) The behaviour of individual, related and groups of South Country Cheviot hill sheep. Animal Behaviour. 11. 507-513.
Study Design Category	2
Assessed by & when	J Bradley 05/02/13

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 upland hill flock. 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: 6 sheep, part of a flock of South Country Cheviot hill sheep, two chosen at random from each of 3 home ranges identified by the shepherd. 9 family groups studied. Not representative of source population.
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 breed of sheep studied were generally representative of eligible population. Unclear if habitat representative as described in another paper.

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: 6 sheep, two chosen at random from each of 3 home ranges identified by the shepherd. 9 family groups all identified by marks, method of selection unclear.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Method explained but other papers not reviewed here required to enable replication. Using an instrument described in Attwood and Hunter (1957) the position of individual marked sheep within the study area was recorded hourly from dawn until dusk on one day a week between September 1958 and March 1959. The same method was used to record the location of each of the family group members during the period 9th Sept 1959 to 16th Aug 1960. The number and location of sheep grazing was also recorded hourly between dawn and dusk between September1956 and September 1959.
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: Range of measurements taken over a combined total of 36 months, no follow up period. All animals had previous experience of grazing areas studied. Location records for July to August 1957-59 were not comparable as due to shearing lambs could not be distinguished from ewes and were therefore counted. Due to the distance from which recording took place the location of each of the marked sheep could not be recorded and it was not always possible to determine if a sheep was grazing. No evidence included re. the effect of supplementary feeding.
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: Lack of evidence concerning supplementary feeding and problems recording activity and differentiating between ewes and lambs may have introduced bias.
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	☐+	Comments: None apparent.

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: Status of population not fully representative of the source population. Breed not fully representative of the source population due to wide range of breeds in source population.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Stock and shepherding practices representative of large areas of UK. Other upland habitats and sheep breeds not considered.

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: Lack of evidence concerning supplementary feeding and problems recording activity and differentiating between ewes and lambs may make the measures unreliable.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐-	Comments: No. Problems recording activity and differentiating between ewes and lambs.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐-	Comments: No. problems recording activity and differentiating between ewes and lambs may have introduced bias Not quantified.
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, Home range behaviour and comparison to family group. The effect of shepherding and supplementary feeding. Seasonal variation in activity.
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-	☐+	Comments: Assessed during 36 months, sufficient to show some significant results. Longer assessment period may be required to assess long term effects.

term effects?		

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Yes.
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted 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.
4.3 Were the estimates of effect size given or calculable?	☐NR	Comments: Home range behaviour and comparison to family group. The effect of shepherding and supplementary feeding. Seasonal variation in activity.
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?	☐-	No statistical analysis other than percentage of location records per grid and sightings recorded on veg types.
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: No.
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 not implemented well, weak replication. A good example of design of an animal behaviour experiment but implementation poor..
5.2 Are the findings generalisable to the		Comments: Due to the wide range of breeds of both

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)?	☐- ☐	sheep within the source population and the poor implementation it would be difficult to extrapolate the results.

Evidence Table

Name of Evidence Review:	Uplands
Name of Review Sub-topic (if any):	Moorland grazing
Review Question	Compare the behaviour of individual sheep and that of the flock as a whole.

Study Details	Authors:	Hunter, RF and Milner, C.
	Year:	1963
	Aim of study:	To test the behaviour of individual, related and groups of South Country Cheviot hill sheep.
	Study design:	2
	Quality Score	+
	External validity:	+
Population and setting	Source population:	UK upland hill flock. Not described in detail
	Eligible Population:	6 sheep, part of a flock of South Country Cheviot hill sheep, two chosen at random from each of 3 home ranges identified by the shepherd. 9 family groups studied. Not representative of source population.

	Inclusion & exclusion criteria:	
	Setting:	Sourhope, Hill Farming Research Organisation, Cheviot Hills, SE Scotland. Vegetation map published in Hunter(1962a)
Methods of allocation to intervention / control	Methods of allocation:	6 sheep, two chosen at random from each of 3 home ranges identified by the shepherd. 9 family groups all identified by marks.
	Intervention description:	Using an instrument described in Attwood and Hunter (1957) the position of individual marked sheep within the study area was recorded hourly from dawn until dusk on one day a week between September 1958 and March 1959. The same method was used to record the location of each of the family group members during the period 9th Sept 1959 to 16th Aug 1960. The number and location of sheep grazing was also recorded hourly between dawn and dusk between September1956 and September 1959.
	Control / comparison description:	No control.
	Sample sizes:	6 individual sheep and 9 family groups comprising of 23 individuals.
	Baseline comparisons:	Location of individual sheep recorded and comparisons made of their home range and those of family members.

	Study sufficiently powered	No power analysis given.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance	Primary outcome measures:	Home range behaviour and comparison to family group. The effect of shepherding and supplementary feeding. Seasonal variation in activity.
	Secondary outcome measures:	
	Follow-up periods:	Assessed over a combined total of 18 months, no follow up period.
	Methods of analysis:	No statistical analysis other than percentage of location records per grid and sightings recorded on veg types.
Results		Individual sheep exhibited home range behaviour which related to the home range groups from which they were selected, but that sheep from the same home range group utilised different areas of that range. Only members from one of the 9 family groups had ranges which were clearly different from each other. Shepherding had little long term effect on sheep behaviour with all individuals returning to their home ranges. Supplementary feeding appeared to have little effect on sheep behaviour. There was a seasonal variation in distribution with animals becoming more dispersed during the period May to October. Grazing activity declined with an increase in daylight hours and the sheep were more active in the first half of the year.
Notes	Limitations identified by	The location records for July to August, Sept 1956-59 were not comparable as due to shearing lambs could not be distinguished from ewes and were therefore counted.

	author:	Due to the distance from which recording took place the location of each of the marked sheep could not be recorded and it was not always possible to determine if a sheep was grazing.
	Limitations identified by review team:	Short study period for each section of the study, no replicates. Small population, one breed. No statistical analysis. No evidence included re. the effect of supplementary feeding.
	Evidence gaps and/pr recommendations for further research:	Longer study period/ study on other upland habitats. Use of different sheep breeds. Statistical analysis of data.
	Sources of funding:	Ministry of Agriculture.

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	Jenkins, D. & Watson, A. 2001. Bird numbers in relation to grazing on a grouse moor from 1957-61. Bird Study, 48, 18-22
Study Design Category	
Assessed by & when	D Martin 18/01/13

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: Moorland bird populations
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: Part of an Estate in E Scotland. Not described in detail but likely to be typical of managed grouse moors. Although different parcels with some intensification between surveys, it is not entirely clear what vegetation types were present at later study
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: Same as eligible population as it is a full farm survey. However some parts don't seem to have been covered in repeat survey

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: Farm-wide survey, so theoretically bias should be low. However methods differed between two surveys. More surveying dine remotely in second period so not covered as intensively as first survey.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Mainly looking at change over time, with grazing cited as main influence. However grazing levels are not recorded in detail, or any surrogate measures. Some broad descriptions of vegetation change, from heather to grass, in some parcels.
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: No real attempt to account for climatic effects etc. Change in burning management also over time. Accounted for to some extent by whole –site approach
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: Variation in methods between study periods. Bird estimates from 1957-61 were most reliable.
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: Basically just bird numbers, no other measures e.g. breeding success or productivity

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:
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: No comparison as such, although unaltered heather moor could be considered as a comparator for those fields that have been converted to grass.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: No analysis as such
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: Really just a comparison of numbers
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:
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)?	☐-	Comments: Limited survey and correlative approach. Different survey methods used between comparison

How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design		periods and slightly different areas included. Although grazing is stated as a main cause of change, it is not really quantified.
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: Not rigorous, or the habitat changes well quantified

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	Jenkins, D. & Watson, A.
	Year	2001
	Aim of study	To investigate changes in bird populations with an increase in grass at the expense of heather moorland, as a result form a shift from grouse shooting towards sheep grazing.
	Study design	3 Observational
	Quality score	-
	External validity	-
Population and setting	Source population	Moorland bird populations. Not well described
	Eligible population	Part of an Estate in E Scotland. Not described in detail but likely to be typical of managed grouse moors. Although different parcels with some intensification between surveys, it is not entirely clear what vegetation types were present at later study
	Inclusion and exclusion criteria	Whole farm study. In second period some of outlying moorland was excluded, probably due to different survey approach, but not clear.

	Setting	Low moorland (250-350m) on farmed grouse moor at Glen Esk, Angus, NE Scotland
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	Prevailing grazing and farming operations
	Control/comparison description	None as such, but there is some comparison between heather dominated moorland, and that which has been transformed to grassland in intervening period.
	Sample sizes	N/A
	Baseline comparisons	N/A
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	All bird species counts
	Secondary outcome measures	N/A
	Follow-up periods	Surveys took place in two periods, one of four years and one of 11 years, covering a 41 year period
	Methods of analysis	No real analysis other than comparisons within and between study periods.
Results		Nine moorland and wading bird species (other than red and black grouse) occurred at high densities (17 pairs per 10km^2 for oystercatcher to 151 pairs per 10 km^2 for meadow pipit) over the four years of the original survey. A further 12 species were found at lower densities with notably redshank and short eared owl increasing 1957 -61.

		In the original survey (1957-61) all parts held golden plover each spring at high density. By 1997 golden plover were not recorded. Where grass has replaced heather since 1961 lapwing, curlew and oystercatcher increased two to four-fold. Numbers of these species were similar in 1997 on heather dominated areas than in the original study period. The change in areas of vegetation is attributed to increased sheep numbers and grassland expansion through subsidies. On a parcel heavily grazed in the later survey years eighteen red grouse, ten black grouse and six grey partridge had been recorded in 1957 and numbers had remained similar through the 1960s. No individuals of any of these species were recorded in the most recent surveys.
Notes	Limitations identified by author	Different survey methods between periods
	Limitations identified by review team	Change in survey approach, no quantification of grazing pressure, no correlative analysis
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	

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

Name of Evidence Review: _____

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

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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?
Study Citation	Jewell, P. L., Güsewell, S., Berry, N. R., Käuferle, D., Kreuzer, M. & Edwards, P. J. (2005) Vegetation patterns maintained by cattle grazing on a degraded mountain pasture. Botanica Helvetica, 115, 109-124.
Study Design Category	2
Assessed by & when	D Martin 19/11/12

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: Alpine mountain pastures with grass and heath communities extending above the natural limit of tree cover (c 1700m). These pastures are declining in use and becoming abandoned by grazing.
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: One site studied, but chosen to be fairly typical of vegetation types and recent pastoral history.
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: Study took place at grazing unit scale – well described in terms of topography, geology, soil and climate, and agricultural history.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: One study site divided in two, with summer cattle grazing regime. Cattle moved from one part to other part way through grazing period. In last year the whole pasture was grazed together, in two periods with a month's gap, and the second period extending into November. Not sure why this was changed. Before the first observation year the pasture had all been run together.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Explanatory variables are measures of grazing behaviour – cattle distribution and activity.
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: No control/ comparison group
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: Single site study – potentially numerous confounding factors of environment and climate. Also a feedback between grazing and vegetation. Feeding of mineral blocks is mentioned, which could affect grazing patterns and diet. Soils were sampled to investigate nutrient status and distribution.
2.5 Is the setting applicable to the UK?	☐+	Comments: Similar vegetation types and grazing livestock (Highland Cattle)

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: Visual estimates in 1998 of percentage cover of species in $1m^2$ quadrats, with weighted ecological indicator values calculated for each quadrat. Seventy quadrats along eight parallel transects at 50m altitude intervals. A further 24 placed subjectively to sample locally occurring vegetation types. Fourteen quadrats randomly selected for biomalss sampling in 1998, and a further 20 in 2000. Also analysed for N and P in three quadrats per vegetation type. Soil samples taken in 82 plots. Grazing observations (location and activity) measured on eight days in 1996 (every 3 hrs) and six days in 1998 (every 2 hrs). Vegetation was also mapped from aerial photographs and some ground truthing – will be subject to a degree of error.

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 grazing only measured in 2 years, and time intervals of observation changed. Soil samples from most, but not all, quadrats.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	Comments: Largely, in relation to the objectives. Short term study will not indicate trajectory of vegetation change. No structural measures or other taxa.
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: Only one study area, no comparison.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: Cattle grazing was re-introduced in 1994, there was no baseline data recorded then, or from first year of cattle observations in 1996. Data collection not designed to detect change. Observational study over a short period.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: Main analysis was of relationships between vegetation type as identified from cluster analysis, and 21 soil variables (7 nutrient or ratio measures from each of three layers). Cattle grazing density in relation to different vegetation types also analysed.
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: Cluster analysis of vegetation samples, and PCA of soils variables. One-way ANOVA of soil variables for each of the four vegetation types identified, with Tukey-Kramer test of significance of pair-wise differences. Grazing intensity expressed in nominal classes (no, light, heavy) and tested among vegetation types with Pearson's Chi square test.

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 presented for Tukey-Kramer tests and results of Pearson's Chi square.
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: Single study site with potential range of confounding factors. Limited livestock observations, with varying grazing regimes in the previous years. Exact numbers of grazing animals not recorded.
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: Provides some limited insight to cattle grazing preferences, on vegetation types broadly transferable to UK situation.

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? g) Do different types of livestock (species and breed), and combinations of livestock, affect moorland habitats differentially?

Study details	Authors	Jewell, P. L., Güsewell, S., Berry, N. R., Käuferle, D., Kreuzer, M. & Edwards, P. J.
	Year	2005
	Aim of study	To determine whether re-introduction of cattle grazing influences vegetation change, in terms of composition and production.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Alpine mountain pastures with grass and heath communities extending above the natural limit of tree cover (c 1700m). These pastures are declining in use and becoming abandoned by grazing.

	Eligible population	One site studied, but chosen to be fairly typical of vegetation types and recent pastoral history
	Inclusion and exclusion criteria	Area chosen based on above.
	Setting	Alpe Nisciora, a 73 ha pasture at 1400m – 1800m on south-eastern flank of Monte Gradiccioli, southern Switzerland.
Methods of allocation to intervention/control	Methods of allocation	One study site divided in two, with summer cattle grazing regime. Cattle moved from one part to other part way through grazing period. In last year the whole pasture was grazed together, in two periods with a month's gap, and the second period extending into November. Not sure why this was changed. Before the first observation year the pasture had all been run together.
	Intervention description	40-60 cattle grazed in summer: end of may to mid July in lower paddock, then to end of September in the upper paddock. In 1988 all run together and two grazing periods ending in November, with one months break. The previous regime from 1994 involved variable numbers grazing over whole area.
	Control/comparison description	NA
	Sample sizes	One study site of 72ha. Vegetation measurements in seventy quadrats along eight parallel transects at 50m altitude intervals. A further 24 placed subjectively to sample locally occurring vegetation types. Fourteen quadrats randomly selected for biomass sampling in 1998, and a further 20 in 2000. Also analysed for N and P in three quadrats per vegetation type. Soil samples taken in 82 plots. Grazing observations (location and activity) measured on eight days in 1996 (every 3 hrs) and six days in 1998 (every 2 hrs).
	Baseline comparisons	No baseline as such. Short term correlative study
	Study sufficiently	No power analysis.

	powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation species composition and spatial distribution of communities. Biomass and soil characteristics of main vegetation types identified.
	Secondary outcome measures	
	Follow-up periods	Cattle grazing was re-introduced in 1994, there was no baseline data recorded then, or from first year of cattle observations in 1996. Data collection not designed to detect change. Observational study over a short period.
	Methods of analysis	Cluster analysis of vegetation samples, and PCA of soils variables. One-way ANOVA of soil variables for each of the four vegetation types identified, with Tukey-Kramer test of significance of pair-wise differences. Grazing intensity expressed in nominal classes (no, light, heavy) and tested among vegetation types with Pearson's Chi square test.
Results		Four main vegetation types were identified with distribution related to topography and pasture management: heathland (*Calluna- Vaccinium*), on remote and steep areas; *Nardus/* heath grassland, which covers about 70% of the area and has similar species to the heat but with *Nardus* dominant; A species-poor (*Carex leporina*) variant of the later that occurs on flat areas where cattle tend to rest; and a variation of *Nardus* grassland with a high proportion of *Agrostis* and *Festuca.* The heat had the highest diversity Whilst the Nardus / heath grassland had the greatest vegetation cover and standing crop, but more than 80% was dead material. The species-poor *Nardus* had a low standing crop and large fraction of living material, suggesting high biomass production. The fourth grassland type is similar in low standing crop and high productivity, and ecological indicators suggest higher fertility. Heath and *Nardus* dominated vegetation were similar in low pH, P concentration, and grater C and N concentration than the other two grasslands. Production in these

		vegetation types are likely to be P-limited, whilst the higher P status of the other two grassland types suggest they are more N-limited. Cattle grazing concentrated in the lower, more fertile areas, so that 40% of the area was grazed only lightly and 50% not at all (mainly heath and Nardus/ heath). Grassland in more fertile areas dominated by *Agrostis, Festuca* or bracken were used most intensely (84% heavily grazed), with most of the species-poor *Nardus* being lightly grazed (71%) and mainly used for resting. There is some evidence of greater movement into less preferred vegetation later in the season, but animals tend to remain near stock buildings and water points, even when little herbage is present. Overall, little change has been observed in the extent of communities over 10 years.
Notes	Limitations identified by author	
	Limitations identified by review team	Limited study with lots of potential confounders from topography and environmental conditions. Positioning of mineral blocks may influence grazing patterns and effects not fully considered.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	ETH Zurich, as part of PRIMALP research collaboration

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? c) What changes have taken place under recent reductions and seasonal changes in sheep grazing, and what is the significance of these changes?

Study details	Authors	Johnston, J (NE)
	Year	2012
	Aim of study	To report the range of stocking rates and condition of a number of SSSIs in the Lake District High Fells SAC
	Study design	3
	Quality score	+
	External validity	+
Population and setting	Source population	Lake District High Fells
	Eligible population	Areas of SSSI under agri-environment agreement
	Inclusion and exclusion criteria	As above – grazing units in AE agreements with stocking rate and CSM data

	Setting	Lake District High Fells
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	All subject to stock reduction – usually to less than annual average of 0.9 ewes per ha from previous ESA rates of typically annual average of 1.3 ewes per ha, or from pre ESA farm stocking rates.
	Control/comparison description	Condition monitoring results under previous heavier grazing regime
	Sample sizes	25 grazing units of 188-3989 ha
	Baseline comparisons	
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Condition assessment, reported for whole unit or component habitats where data allows/ appropriate
	Secondary outcome measures	
	Follow-up periods	Variable. Agreements started between 1999 and 2010, with most around 2003.
	Methods of analysis	Comparisons
Results		Annual average stocking rates range from 0.3-0.89 ewes per ha, with most in the range 0.5-0.69. Most units reported to be in recovering or good condition compared with previous assessments. Localised issues are reported, for example the recovery of ledge and tall herb communities being confined to less accessible areas, and montane heath

		not showing signs of recovery in some locations. In general recovery of these habitats appears slowest at the higher end of the range of stocking rates. Blanket bog recovery is reported in sites where the annual average rate is around 0.4 ewes per ha, and up to 0.67 ewes per ha in summer. Grass dominated units allow for higher stocking rates, which may impact on areas of sensitive habitat. The lower annual rates tend to be associated with off-wintering. Timescales of recovery varies, but suppressed and fragmented heathland is shown to respond quickly to reductions in grazing pressure. The initial response of unpalatable grasses such a mat grass is to increase in structure, leading to concerns of 'undergrazing'. The longer term response may be an alteration in the competitive balance of palatable and unpalatable species. The lower stocking regimes promote patchiness of grazing with sheep exploiting more palatable vegetation and less likely to graze less preferred communities. An exercise to collate stocking rate and condition monitoring data from a number of grazing units in the Lake District high fells indicated that annual average stocking rates, largely of sheep, has resulted in improvement in habitat condition over a period of around10 years. Blanket bog has responded well where annual average rates are below 0.4 ewes per hectare, often through off-wintering. Response of other sensitive habitats such as ledge and montane communities is more variable. These are often small areas within grassy fells which can otherwise accommodate higher stocking rates. Stock reduction tends to promote patchy grazing, with avoidance of less palatable species.
Notes	Limitations identified by author	Time available for study and analysis. A limited piece of work to collate readily available data, and does not include more detailed monitoring data available for some sites.
	Limitations identified by review team	Casual/ observational study with subjectivity. But useful collation of information for a number of sites.
	Evidence gaps and/pr recommendations for	

	further research	
	Sources of funding	NE internal

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? c) What changes have taken place under recent reductions and seasonal changes in sheep grazing, and what is the significance of these changes?
Study Citation	Johnston, J. (2012) Stocking rates and Condition Assessment on Sites of Special Scientific Interest (SSSIs) within the Lake District High Fells Special Area of Conservation (SAC).
Study Design Category	3
Assessed by & when	D Martin

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: Would benefit from more formal monitoring and analysis, but has not been resourced.
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments:
1.3 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used?	☐ Defensible	Comments: Casual collation and summary of agri-environment site information. Monitoring is generally CSM, a structured largely visual assessment, rather than quantitative monitoring. Site cover significant area of Lake District Fells (c 30 000ha)

- Is the selection of cases / sampling strategy theoretically justified?		

Section 2: Study Design		
2.1 How defensible / rigorous is the research design / methodology? For example: -Is the design appropriate to the research question? -Is a rationale given for using a qualitative approach? - are there clear accounts of the rationale for sampling, data collection and data analysis techniques used? - Is the selection of cases / sampling strategy theoretically justified?	☐ Not Sure N/A	Comments: Not really a designed study but summary of a range of cases. Full coverage of SAC area

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?	☐ Not Sure / inadequately reported	Comments: Collation of stocking rate and CSM data for each site.

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	Comments: NE Conservation Adviser with long history of working on these sites
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?	☐ Not Sure	Comments: Carried out under advisers own initiative. A data collation exercise, involving some calculation of average stocking rates. Not sure if checked by other observers.

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 N/A	Comments: No analysis as such but reporting of a range of stocking rates and within-site comparisons with CSM results
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 is clear. Sites cover a range of habitat types and scenarios
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?	N/A	Comments:
5.4 Are findings convincing? For example: -findings clearly presented? -finding internally coherent? -Extracts from original data included? -data appropriately referenced? -reporting clear and coherent?	☐ Convincing	Comments:
5.5 Are the findings relevant to the aims of the study?	☐ Relevant	Comments:

5.6 Conclusions For example: -how clear are the links between data interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?	☐ Not sure	Comments: Degree of subjectivity involved, as limited numerical analysis. The observations are fairly reliable

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?	N/A	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? - Was the data collection and record keeping systematic?	☐ +	Comments: Not a rigorous quantitative analysis but a useful collation of data from a large number of grazing units that would benefit from more rigorous monitoring and analysis.