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	Rawes, M.
	Year	1983
	Aim of study	To investigate the response of the vegetation of two bogs when sheep were excluded.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	High level Eriophorum dominated blanket bog well described with reference to previous studies
	Eligible population	The study area is at the higher end of the altitudinal range of the habitat
	Inclusion and exclusion criteria	Chosen to be fairly representative of surrounding area, but subjective and small plots. The two areas differ in aspect, slope, depth of peat and surrounding vegetation, but both *Eriophorum* dominated with some *Calluna* and *Sphagnum*.
	Setting	Troutbeck Head and Silverband, Moor House NNR, on the east and west sides of the

		Pennine summit ridge near Cross Fell. Both areas at 685m asl, on peat 1.5-2m deep
Methods of allocation to intervention/control	Methods of allocation	Not stated, likely to have been chosen subjectively to represent the predominant bog vegetation and conditions at the site.
	Intervention description	Treatment is long-term removal of grazing. Small plots, un-replicated. Subjectivity in sample selection.
	Control/comparison description	Open plot subject to prevailing agricultural grazing levels.
	Sample sizes	Two exclosures. Vegetation sampled from 500 point quadrats, on 5 transects. Whole plot mapped, and detailed plant locations in 3 25x25 cm quadrats per plot.
	Baseline comparisons	Measurements made in first year of exclosure.
	Study sufficiently powered	No
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Cover and height of each species, extent of vegetation stands. Measures of density and vertical structure from pin hits at different heights above ground.
	Secondary outcome measures	
	Follow-up periods	Exclosures in place for 14 years
	Methods of analysis	Fairly basic – binomial sign test of species change. No statistical comparison with grazed plots.
Results		At Troutbeck, heather increased on average by 14% each year, following grazing exclusion. Cover was not recorded at Silverband in the first year of exclosure, but

		spread rapidly in the last 6 years to reach 2.4%. Heather cover decreased significantly from 6% to 3% in the Troutbeck grazed plot, but was not recorded at Silverband. There were no other significant changes in either grazed plot, other than an increase in *Deschampsia flexuosa* at Troutbeck. In both exclosures there was an increase in *Empetrum nigrum* and *Rubus chamaemorus*. Eriophorum vaginatum declined in both, significantly so at Troutbeck, but remained the most abundant species. *Narthecium ossifragum* was present at Silverband only, and doubled in cover here. Vegetation density increased by 20% at Troutbeck and 40% at Silverband, with increases generally above 10cm and decreases below. E vaginatum became markedly less common below 10cm. Plot mapping at 1:50 scale shows generally similar trends in cover of key species, with some differences, for example low cover of *E nigrum* in 1980 compared to point quadrat data with early increases in cover replaced by *C vulgaris*. Bare peat generally reduced in cover, but some erosion continued. *Sphagnum* spp initially flourished, but reduced with Calluna expansion and drying effects. At Silverband, *N ossifragum, Carex* nigra and *Eriophorum* spp colonised bare peat. Micro-habitat quadrats showed that movement and expansion of species was largely rhizomatous; seedlings were never found. Two quadrats on *Sphagnum* hummocks showed heather becoming dominant, then gaps appearing in the canopy allowing some *Sphagnum* recovery. In a bog asphodel patch this species changed from a short-grazed turf to a thick mat of dead leaves, eliminating a co-dominant liverwort. In eroded bare peat vascular plant colonisation increased. Recovery was due to lack of grazing in grazing-sensitive species, but also through lack of trampling on species such as *E nigrum* which are usually avoided by grazing sheep. Lack of grazing leads to change in size, frequency and position of species present rather than invasion by new species. Whilst some species such as *Calluna* and *N ossifragum* locally increased in dominance, *Trichophorum caespitosum* patches became less pronounced. Changes in moisture regimes as a result f increased plant growth led to increased frost

		damage of heather. Surface water contributed to the slow colonisation of bare peat.
Notes	Limitations identified by author	
	Limitations identified by review team	Lack of replication, limited analysis
	Evidence gaps and/pr recommendations for further research	Larger scale experiments to investigate agricultural benefits of grazing regimes that sustain greater vegetation cover and populations of otherwise preferentially grazed species.
	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? h) What are the effects of absence or removal of grazing on moorland biodiversity and other ecosystem services?

Study details	Authors	Rawes, M & Hobbs, R
	Year	1979
	Aim of study	Effects of burning, sheep grazing removal and increased grazing on the condition of blanket bog vegetation.
	Study design	2 comparisons, but with some replication in aspects of the study.
	Quality score	+
	External validity	+
Population and setting	Source population	Source population is high level blanket bog in the North Pennines. Vegetation and topography of the bog on the site (Moor House NNR) is quite well described.
	Eligible population	Not entirely clear but largely same as source population so likely to be representative.
	Inclusion and exclusion criteria	Selection not described, and no indication on representativeness at baseline. Twelve plots are reported, of varying size, the largest 3 ha, plus experimental blocks for burning

		and grazing experiments. Selection not described but presumably chosen to be fairly representative.
	Setting	Moor House NNR, North Pennines. 540-606m
Methods of allocation to intervention/control	Methods of allocation	Not described for grazing exclusion plots. Burning experiment on four blocks to cover altitudinal variation, with burning treatments randomised.
	Intervention description	Grazing removal (long-term) on 12 plots, with factorially applied grazed and ungrazed and ten and twenty year burning rotations. Grazing experiment has a light, heavy and ungrazed treatment.
	Control/comparison description	Stated as ungrazed, un burned, but lightly grazed plots are more typical of practice.
	Sample sizes	12 ungrazed plots, four replicates of burning/ grazing treatment combinations. Pin hit cover measures made at 25-40 frames, x ten pin hits.
	Baseline comparisons	N0 baseline measurements for main study (data given for 7 and 18 year differences between grazed and ungrazed. The grazing experiment has baseline year vegetation data.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Species frequency and cover comparisons.
	Secondary outcome measures	Age distribution of heather in burned areas. In the burning experiment a range of other measures included seedling no, number of shoots etc in different treatments.

	Follow-up periods	Grazing removal data reported for 7 and 18 years. Burning experiment 22 years. Data reported for each of 7 years in grazing experiment.
	Methods of analysis	Mean values and t-tests. Some detailed mapping comparisons of quadrats.
Results		Following exclusion of sheep there was a marked change to the vegetation of *Eriophorum vaginatum* dominated blanket bog in the North Pennines. There were increases in *Calluna* at both the 7 year and 18 year timeframes and a corresponding decline in *E. vaginatum*. The increase of *Calluna* at 18 years was not significant but the continued decline of *E. vaginatum* was. Lichen response was also marked with increases in both cover and biomass. The mapping work focussed on the wettest blanket bog. This also showed increases in *Calluna* but it is acknowledged that climatic factors could account for this and there was poor control. It was concluded that it is clear that sheep grazing has a major influence in determining the botanical composition of blanket bogs.
Notes	Limitations identified by author	Limited range of stocking rates, low level of agricultural grazing in the comparisons.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	How long bog vegetation takes to return to 'steady-state' after major disturbance
	Sources of funding	

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

Name of Evidence Review: ___Upland_____

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

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Rawes, M & Hobbs, R. 1979. Management of semi-natural blanket bog in the North Pennines. Journal of Ecology, 67, 789-807
Study Design Category	2
Assessed by & when	D Martin

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Comments: Source population is high level blanket bog in the North Pennines. Vegetation and topography of the bog on the site (Moor House NNR) is quite well described.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐+	Comments: Not entirely clear but largely same as source population so likely to be representative.
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, and no indication on representativeness at baseline. Twelve plots are reported, of varying size, the largest 3 ha.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐NR	Comments: Not clear how any selection bias was avoided.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Two experiments are reported, covering sheep grazing vs no grazing and two burn rotations and unburned. Grazing reflects the prevailing levels and heavier grazing. The paper also reports the effect of grazing exclusion on vegetation, and by inference the results of on-going agricultural grazing levels (although they are low compared with other upland areas).
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐+	Comments: No contamination reported.
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: Four replicates in the burning trial would have removed some of the effects of topography and variation in the vegetation composition. No replication in grazing trials however. For the grazed and ungrazed comparisons direct measurements are made in adjacent areas – similar vegetation at start of the study.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes, representative of higher level extensive bog. Typical vegetation communites of this habitat.

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: Measurements in the grazing exclusion study involved pin-frame assessments of cover – expressed as mean number of contacts per ten pin frame. Thirty or forty frames per plot. Similar technique for the grazing study, but in the first year of the burning study cover was estimated on the domin scale. Vegetation structure was also measured by contacts at different height zones. Phenology including number of shoots, flowering heads etc measured.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	☐++	Comments: Yes

3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments: Yes, in terms of the objectives. Basically measures of vegetation community composition.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Yes
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: Sheep removal for 20+ years, although varies between plots. Grazing experiment assessed after 8 years.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Yes

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: Mainly just single variables, either grazing level or burning frequency.
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: basic t-tests and comparisons of mean figures.
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 t-tests.
Section 5: Summary		
5.1 Are the results of the study internally		Comments: Grazing exclusion in a reasonably large

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	☐+	number of plots, and over long timescale. Burning experiment replicated. However the grazing experiment is un-replicated and includes a very limited range of grazing levels, which may not be very representative of elsewhere.
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: Broadly, due to number of plots and duration etc. Grazing levels may not be typical of many areas.

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	Rawes, M. (1983) Change in two high altitude blanket bogs after the cessation of sheep grazing. Journal of Ecology, 71, 219-235
Study Design Category	2
Assessed by & when	D Martin 28/11/12

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: High level Eriophorum dominated blanket bog well described with reference to previous studies.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: Yes, although the variant in the study area is at the higher end of the altitudinal range of the habitat.
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: Chosen to be fairly representative of surrounding area, but subjective and small plots. The two areas differ in aspect, slope, depth of peat and surrounding vegetation, but both *Eriophorum* dominated with some *Calluna* and *Sphagnum*.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: Treatment is long-term removal of grazing. Small plots, un-replicated. Subjectivity in sample selection.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Grazing vs no 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:
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. The plots vary in topography peat depth and wetness. Grazing history likely to have been slightly different. Grazing by red grouse and voles were significant.
2.5 Is the setting applicable to the UK?	☐++	Comments:

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Vegetation mapped at 1:50 based on a fixed grid, with point quadrat samples at fixed intersections. Measurements at different canopy heights. The 1:50 maps do not always agree with point quadrat results in terms of cover of key species. Three small fixed quadrats also established in each treatment, mapped at 1:1, to chart change over time in different microabitats.
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: Vegetation measurements made four times over the 14 years covered. Non-vascular plants only on two occasions at beginning and end of this period.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+	Comments: In terms of vegetation composition and structure. Change in grazed plots not presented.

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: 14 years grazing exclusion

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: Only two exclosure and grazed plots, in different locations.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: Only grazing or no grazing
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: Fairly basic – binomial sign test of species change. No statistical comparison with grazed plots.
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 change in percent cover of species from point quadrat measurements.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential	☐-	Comments: No replication, limited analysis or consideration of a range of factors. Limited presentation of data from grazed plots.

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

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	Roberts, J
	Year	2002, 2003
	Aim of study	To investigate the effects of large-scale stock removal on flowering of uncommon and rare arctic/ alpine plants on the Cross Fell Massif, N Pennines
	Study design	3
	Quality score	+
	External validity	+
Population and setting	Source population	Upland and montane habitats of N Pennies
	Eligible population	As above, in areas where summer grazing had been removed.
	Inclusion and exclusion criteria	Focus tended to be on areas known or likely to hold unusual species – e.g. flushes. Limestone outcrops

	Setting	The Massif around Cross Fell and Great Dun Fell in the North Pennines, on the Cumbria/ Durham border. Generally above 600m.
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	Removal of summer sheep grazing following the 2001 foot and mouth outbreak
	Control/comparison description	Known location and extent of arctic/ alpine species from botanical records
	Sample sizes	Not known
	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	Records of presence of a wide range of species, particularly key montane and arctic/ alpine spp
	Secondary outcome measures	Degree of flowering – number of flowering heads of target spp
	Follow-up periods	One year – fells visited
	Methods of analysis	N/A
Results		In the season following a lack of summer grazing due to foot and mouth livestock movement restrictions and culls, and before the fells were restocked, the compliment of species present in various habitats were more apparent due to increased productivity and flowering. Of greatest significance was the finding of sheathed sedge (*Carex vaginata*) at two locations, and profuse flowering of colonies of alpine foxtail (*Alopecurus borealis*) new and known sites. Significantly greater flowering was

		recorded for known rare species including alpine forget-me-not (*Myosotis alpestris*) and marsh saxifrage (*Saxifraga hirculus*) New tetrad records and/ or new upper altitudinal records for Cumbria and in some cases for the British Isles were found for a significant number of species. Extensive flowering was observed of typical limestone short turf species, including mountain pansy (*Viola lutea*), wild thyme (*Thymus polytrichus*) and spring sandwort (*Minuartia verna*). In the relatively low diversity moorland and blanket bog habitat the increased flowering allowed subtle and complex variations in plant communities, not usually so visually apparent under typical grazing, to be readily observed. A year of no grazing over an extensive upland area allowed greatly increased phenological expression of the full range of species present, resulting in records for a species previously unknown in England (sheathed sedge) and new sites identified for a wide range of species of varying rarity, including alpine foxtail. Profuse flowering was observed in many other species which rarely or sparsely flower under typical grazing. Increased flowering allowed subtle variations in composition of relatively low diversity moorland habitats to be observed. Re-introduction of grazing in the following year saw reduced flowering of species such as marsh foxtail, in terms of frequency and size of flowering stems, and length of flowering period, compared to the ungrazed season. It is noted though in Roberts (2010) that some grazing is required to maintain open conditions for poor competitors, as noted with marsh saxifrage where grazing has been excluded for 10 years.
Notes	Limitations identified by author	N/A
	Limitations identified by review team	Opportunistic casual surveys rather than a planned systematic survey. At least some of the new records may be down to the increased survey effort rather than the effects of stock removal *per se*.
	Evidence gaps and/pr	Monitoring of longer term effects of changes in grazing pressure and plant responses.

	recommendations for further research	
	Sources of funding	N/A

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? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Roberts, J. (2002) After foot and mouth, Cross Fell in bloom. The Carlisle Naturalist, Vol10, No 2, pp. Roberts, J. (2003) Cross Fell Update, 2003, Volume 11, no. 2, pp. 47-52. Carlisle Naturalist
Study Design Category	3
Assessed by & when	D Martin 27/12/12

Section 1: Theoretical approach		
1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question?	☐ Appropriate	Comments: Citations are collations of records from semi-systematic but opportunistic surveys of botanically rich areas on Cross fell massif in North Pennines following sheep removal during the 2001 Foot and Mouth outbreak
1.2 **Is the study clear in what it seeks to do?** For example: - is the purpose of the study discussed – aims/objectives/research questions? -is there adequate / appropriate reference to literature? - are underpinning values / assumptions discussed?	☐ Clear	Comments: Yes – to investigate the response of plants to a season of no grazing – with a particular interest in finding new records for species in the locations visited.
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?	☐ Not Sure	Comments: Not really a piece of research/ scientific study as such, but adds to the body of knowledge of species distribution.

- 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?	☐ Defensible	Comments: Not intended as a research study, but survey/ collation exercise

Section 3: Data Collection		
3.1 How well was the data collection carried out? For example: -Are data collection methods clearly described? -Were the appropriate data collected to address the research question? - Was the data collection and record keeping systematic?	☐ Appropriately	Comments: Records of presence with location data. Areas likely to hold unusual/ rare species targeted. Search approach rather than sampling methods.

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: Committed amateur botanists
4.2 Is the context clearly described? For example - were observations made in a sufficient variaty of circumstances? - was context bias considered?	☐Clear	Comments: The circumstances are well described
4.3 Were the methods reliable?	☐ Reliable	Comments: Not designed as a co-ordinated scientific study, but casual records

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?	Nor really relevant	

Section 5: Analyses		
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Rigorous	Comments: No analysis as such, but records collated and fed into mapping, botanical flora etc.
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: As well described as possible. Context and significance clear
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: No analysis as such
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: Rare opportunity to examine effects on a range of vegetation types and phonological response of species to large-scale grazing removal.
5.6 Conclusions For example: -how clear are the links between data	☐ Adequate	Comments: Largely circumstantial, but the area is well studied botanically so whilst there is some confounding of findings through

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?		probable increased survey effort, the effect is significant enough to conclude that there is a link with grazing removal.

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

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: Findings undoubtedly reliable, but not a systematic or repeatable 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?

Study details	Authors	Ross, S
	Year	2000
	Aim of study	To investigate the role of summer grazing on heather stands burned at different ages, in terms of *Molinia* control and heather regeneration.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Wet heath dominated by dwarf shrub, with Purple moor-grass (*Molinia*)
	Eligible population	Moorland with small-scale mosaic of both molinia- dominated and dwarf-shrub dominated vegetation types.
	Inclusion and exclusion criteria	Desired vegetation types and mosaics.
	Setting	Redesdale Research Farm, Northumberland

Methods of allocation to intervention/control	Methods of allocation	Likely to have been subjectively chosen.
	Intervention description	Plots established on 2 burned areas, with two levels of grazing based on target off-take. Ungrazed controls.
	Control/comparison description	Ungrazed plots in both
	Sample sizes	No replication. Quadrat number not reported
	Baseline comparisons	Likely to have been similar – burned at same time. Vegetation measurements made in summer after burning.
	Study sufficiently powered	Not reported. Un-replicated so power low.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation frequency measures, particularly Molinia and Calluna
	Secondary outcome measures	none
	Follow-up periods	Measured over 3 years
	Methods of analysis	Not reported
Results		Molinia cover decreased by at least 50% under both grazing regimes for heather burned at 8 years. Heather increased in cover under low grazing, but little under high grazing. For heather burned at 22 years, Molinia cover increased by 50% in the ungrazed control, but decreased under both grazing treatments. Heather cover decreased under high grazing, but increased slightly under low grazing and control.

		No regeneration data presented for heather burned at 8 years. For 22 years, seedlings increased in the first autumn to 50 seedlings per m^2, then declined over the rest of the monitoring period to half of this. Similar initial regeneration rates were sustained in second and third year under low grazing. Regeneration under high grazing was low (<10 seedlings per m^2). Summer only grazing appears to be effective in controlling Molinia after burning. The lower grazing rate (equivalent to approximately 33% utilisation) was effective in reducing Molinia cover, whilst allowing regeneration. The higher rate compromised regeneration.
Notes	Limitations identified by author	
	Limitations identified by review team	Lack of replication, analysis not presented
	Evidence gaps and/pr recommendations for further research	Application of this type of summer only grazing regime at a larger scale to explore the usefulness for controlling molinia and enhancing heather regeneration under burning.
	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	Ross, S. (2000) Molinia management using sheep grazing preferences. In: Molinia management in ESAs and the uplands. ADAS workshop 14-15 June 2000
Study Design Category	2
Assessed by & when	D Martin 14/11/12

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 moorland well described in other parts of the report.
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: moorland with small-scale mosaic of both molinia- dominated and dwarf-shrub dominated vegetation types.
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: Heather stands of different ages with Molinia present. No indication of how selected, but two different ages of heather 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?	☐NR	Comments: Likely to have been subjectively chosen
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Plots established on 2 burned areas, with two levels of grazing based on target off-take. Ungrazed controls.
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 4 years. This is adequate for effects on established Molinia (Grant et al 1996)
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: Management is by put and take to achieve desired off-take levels – may result in unintended variation from target.
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: Likely to be representative of wet moorlands with Molinia
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The 8 year burn is fairly typical of intensive grouse management. The 22 yr rotation longer than usual. Grazing levels may be similar to some lower and higher rates experienced on this type

		of moorland

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: Methods not given in detail – quadrat (how many, where?) measurements each year of vegetation re-establishment including Calluna and Molinia relative frequency, and Calluna regeneration.
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: Calluna regen in 8 year old burned heather stand not presented.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Yes, fairly simple 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)?	☐++	Comments: Yes
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. Will only measure initial effects. Would be good to follow for longer

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: Likely to be – burned at same time

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?	☐+	Comments: Some indication from previous study (Grant et al) of effects of the imposed grazing levels. This older study was however on well developed tussocks.
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	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	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: Little information given on analysis methods. No replication. Some data not presented. This is not a peer reviewed paper – there may be another source for this 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:

Grant, S.A., Torvell, L, Common, T.G., Sim, E.M. & Small, J.L. (1996). 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

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	Rushton, S.P., Sanderson, R.A., Wildig, J. &Byrne, J.P.
	Year	1996
	Aim of study	To demonstrate how the results of small-plot, field and farm-scale experiments can be used in combination with modelling to make long-term predictions of the effects of change in grazing pressure on vegetation communities.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	The source population is the extent of grazed semi-natural upland plant communities. Not presented in great detail
	Eligible population	Two upland farm units Mid Wales, Northumberland) with typical open hill vegetation communities chosen. The approach is pragmatic as the two areas are experimental hill

		farms
	Inclusion and exclusion criteria	Moorland vegetation communities
	Setting	Research farms: Pwllpeiran, Cambrian Mountains, mid-Wales and Redesdale, Northumberland.
Methods of allocation to intervention/control	Methods of allocation	Not clear – partly opportunistic – research farm used.
	Intervention description	Three experiments – plot, field and farm-scale. ESA stocking rate, ESA-30% and ungrazed treatments at plot scale (2ha), and first two treatments at field-scale (20ha). Farm scale experiments (117 ha are split in two) were grazed at 2.1 ewes per ha (to reflect typical sheep grazing) and 1.48 ewes per ha.
	Control/comparison description	At farm scale 'typical' provides a comparison. No clear control at other scales
	Sample sizes	Treatments not replicated, so one plot of each. Vegetation assessed in fixed quadrats – 15 per treatment at plot scale, 30 at field and 166 at farm.
	Baseline comparisons	Vegetation measures and derivation of NVC communities represented at start of experiment.
	Study sufficiently powered	No – no calculation, no replication, small sample.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	Vegetation composition within fixed quadrats. NVC derived from ordination.
	Secondary outcome measures	None

significance)	Follow-up periods	Treatments in place for five years (1990- 1994)
	Methods of analysis	Main analysis is ordination of quadrats to identify closes NVC community. Change only presented in terms of derived NVC types, so magnitude of change in terms of cover and frequency of species, species gain and loss etc not known. Paper largely focuses on a Markovian modelling approach to Predict vegetation change, based on the recorded change 1990-1992, and compared against the recorded vegetation in 1994.
Results		Vegetation dynamics appeared to be linked to scale, with more dynamic change in the smaller plots than in farm scale experiment. A marked change in the proportion of quadrats in vegetation type was observed at this scale. The model predicts increases in the dwarf shrub communities over 10 years, although there is a poor fit between observed and predicted after 5 years. At the field scale there was a predicted and observed increase in heather/ bilberry community after 5 years in the lower grazing treatment (0.83 ewes per ha, Apr- Oct). Change appeared to be slow at the farm-scale plots, and most communities are expected to persist under the lower stocking rate (1.48 ewes per ha). A decline in Nardus was predicted. The accuracy of the model predictions increased with the size of the experimental unit, being greatest at the farm scale, where change was slowest. Simplistic management prescriptions may not take adequate account of the ecological processes affecting vegetation at different scales.
Notes	Limitations identified by author	Caution needed in predicting change using NVC community predictions, due to variability in frequencies of key species within a community. Large-scale plots would benefit from more fixed quadrats than in this experiment coupled with monitoring of sheep behaviour to identify different grazing pressures through the area.
	Limitations identified by review team	Lack of replication, lack of detailed analysis of actual species response. Often poor fit between model and observed response. The treatments are a bit ambiguous in terms of timing (all summer only, or are farm-scale applied year round?) Stocking rates at farm-scale higher than other two scales, so confounds scale effect.

	Evidence gaps and/pr recommendations for further research	Little presented – an acknowledgement that larger scale grazing experiments need to take account of spatial affects and heterogeneity of grazing pressure, as mentioned above.
	Sources of funding	MAFF, as part of the ADAS Hills and Uplands project.

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	Rushton, S.P., Sanderson, R.A., Wildig, J. &Byrne, J.P. (1996). The effects of grazing management on moorland vegetation: a comparison of farm unit, grazing paddock and plot experiments using a community modelling approach. Aspects of Applied Biology 44, 211-219
Study Design Category	2
Assessed by & when	D Martin 22/10/12

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: The source population is the extent of grazed semi-natural upland plant communities. Not presented in great 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: Two upland farm units Mid Wales, Northumberland) with typical open hill vegetation communities chosen. The approach is pragmatic as the two areas are experimental hill farms. Broad vegetation, altitude and rainfall are presented. These however cannot represent the range of upland communities and topography through the geographical range of English and Welsh uplands.
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: It is not clear how the small plot or field-scale areas were assigned. Rationale for choosing three scales and the size of each set of study areas not entirely clear, and whether they can really be considered 'field' and 'farm' scale. Is there some redundancy, especially between 2ha plot and field scale measurements?

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐-	Comments: It is not clear how three grazing treatments were assigned, and appears to be no replication. It is possible that there will be confounding factors through variation in soils, vegetation productivity etc. It is not clear how the two vegetation types mentioned at Pwllpeiran are apportioned to treatments.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Described in terms of stocking rate imposed. Chosen to replicate typical ESA stocking rates, aimed as habitat maintenance, and reduced rate, which may facilitate vegetation restoration – the reduction figure is arbitrary. Could be replicated, but there may be technical issues with achieving the stocking rates on the small plots, not documented here. Question of validity of different stocking rates at farm-scale – will confound investigation of the scale effects investigated?
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 each year 1990-1994. Implementation consistent across experiments and sites. Longer exposure would allow further vegetation change to be detected, but has to be balanced against other practical considerations, and need for reportable findings.
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 apparent.
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: Vegetation described is typical of some of the most widespread upland communities, but only two sites so will not be fully representative of the geographical range. The NVC studies describe the

		vegetation present – generally common and widespread communities.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐-	Comments: The interventions included a 'typical' stocking rate at the farm scale site, with arbitrary 30% reduction in the other treatment, not sure of the rationale for this as it is above the ESA rate investigated in the other experiments. The 30% reduction on ESA rates is fairly arbitrary.

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: Outcome measures are vegetation communities within fixed quadrats. The method of vegetation recording is not explained – presumably each species with estimate of cover, and frequency, but not clear. A 'pseudo-quadrat' approach used to derive an ordination and assign each field quadrat to a NVC type. Appears to be large proportion of quadrats change over time e.g. ESA+30% plot scale – big increase in Nardus dominated grass- was this a real change or artefact of the method?
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: It would appear so from the paper.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Only vegetation composition assessed, but this is in line with aims.
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: Veg composition is the main outcome, and change in this over time. However outcomes reported are largely those of the model and difficult to get a feel for the real change.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: All treatments in place for same length of time
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR	Comments: Treatments in place for 5 years – has allowed changes in vegetation to take place. All vegetation management studies benefit from long-term experiments, but has to be balanced against other needs.

	☐NA	

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: Whist broadly similar in vegetation communities present, proportions seen to vary – e.g. Nardus much higher in ESA plot, Agrostis/festuca in ungrazed plot at outset. Some difference in proportions of heathland communities at field-scale, Farm scale plots reasonably similar.
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 analyses, but no replication, likely to be low-powered.
4.3 Were the estimates of effect size given or calculable?	☐+	Comments: Effect size apparent in terms of proportion of quadrats changing communities, both observed in 1994 and predicted from model.
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: Main analysis is ordination of quadrats to identify closes NVC community. Would seem to be appropriate, as an objective method. Proportions of quadrats in each community at baseline and 1994 presented for each treatment. Change only presented in terms of derived NVC types, so magnitude of change in terms of cover and frequency of species, species gain and loss etc not known. Discussion makes the point that increase in one community at Redesdale is largely down to one species which is not of great conservation benefit (M caerulea)
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: Statistical testing not possible in two of the experiments due to low cell counts
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential	☐-	Comments: Although based on field measurement, largely a modelling paper and more could be made of real change in species frequency and composition at different scales. Changes presented as proportion of quadrats assigned to NVC types, but little indication

confounders)? Were there any significant flaws in the study design?		of significance of underlying vegetation change. Lack of replication, and indication of differences between plots (small and field-scale) at baseline. Validity of comparing Farm-scale experiments with other two is questioned – higher stocking rates involved.
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 broadly typical of main upland vegetation types, with conservation grazing levels included in some of the treatments. Little transferable findings in terms of effects of stocking rates, but highlights that over large areas grazing pressure is more heterogeneous, affecting different vegetation types differently.

Evidence Matters Newsletter Issue 10**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	Sibbald A M
	Year	Published 2008. Work done 2001
	Aim of study	Effects of social behaviour on the spatial distribution of sheep grazing a complex vegetation mosaic
	Study design	RCT
	Quality score	=QA5.1 Each plot had 250 scan samples of positions and behaviours of individual sheep carried out during a 2-week observation period – 25 scans between 7:30 – 21:30/day for 11 days (extra day to account for low cloud obscuring plots)
	External validity	=QA5.2 sufficient details given but not necessarily 100% transferable from Scottish Blackface to English breeds of sheep
Population and setting	Source population	Glensaugh Research Station, Northeast Scotland. 6 hill plots 100mx100m on a hillside facing NNW. Highly fragmented mosaic, predominately Calluna with numerous patches of Agrostis-Festuca grass. 36 yearling female Scottish Blackfaces from a single flock used.
	Eligible population	Highly fragmented mosaic, predominately Calluna with numerous patches of Agrostis/Festuca grass.
	Inclusion and exclusion criteria	

	Setting	
Methods of allocation to intervention/control	Methods of allocation	36 yearling female Scottish Blackfaces from a single flock used. Allocated at random into 6x6 groups, individually marked for distance id. Each group allocated to one plot.
	Intervention description	
	Control/comparison description	
	Sample sizes	36 yearling female Scottish Blackfaces from a single flock used. Allocated at random into 6x6 groups, individually marked for distance id. Each group allocated to one plot. Each plot had 250 scan samples of positions and behaviours of individual sheep carried out during a 2-week observation period – 25 scans between 7:30 – 21:30/day for 11 days (extra day to account for low cloud obscuring plots)
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Scanned observations by telescope from approx.500m marked directly onto vegetation maps and transferred to GIS maps of the plots
	Secondary outcome measures	Behaviour recorded as grazing, lying, standing, walking with head up or drinking.'
	Follow-up periods	
	Methods of analysis	Comparing number of sheep locations on individual patches with expected values based

		on the areas of the patches
Results		Sheep spent 69% of their time grazing, 26% lying down, remainder either walking or standing, drinking<0.1%. When grazing, 69% was on grass patches. Sheep spent more time than expected on larger patches (highly significant on all plots) 'When observations of sheep grazing on the most preferred grass patch in each plot were analysed, there were many more occasions when 4,5 or 6 sheep grazed there together than would have been expected from the frequency with which sheep visited those patches' 'Sheep preferred to graze on some of the largest grass patches in the mosaic' 'Frequencies with which sheep were seen in groups were significantly higher than would be expected simply from the number of times that individual sheep visited those particular patches, suggesting that the animals made positive choices to graze there together' 'Patch sharing was necessary for groups to maintain their (4.9m) preferred spacing while grazing grass'
Notes	Limitations identified by author	
	Limitations identified by review team	Short observation period
	Evidence gaps and/pr recommendations for further research	'Both experiments' (this one and Hester et al (1999)) 'used groups of 6 Scottish Blackface sheep and further research will be necessary to determine the extent to which the results can be generalised to larger groups, more sociable breeds of sheep and different degrees of environmental heterogeneity'

	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): _____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 social behaviour on the spatial distribution of sheep grazing a complex vegetation mosaic. Sibbald A M. Applied Animal Behaviour Science 115 (2008) pp 149-159
Study Design Category	1 2
Assessed by & when	Alison Hiles 8/2/2013

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: Glensaugh Research Station, Northeast Scotland. 6 hill plots 100mx100m on a hillside facing NNW. Highly fragmented mosaic, predominately Calluna with numerous patches of Agrostis-Festuca grass. Plots fenced in 1990 with regular summer grazing until 2001 when this experiment began. This included grazing studies in 1998 and 1999. Prior to this experiment the plots were grazed intermittently in winter. All sheep removed in April 2001
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: Vegetation maps produced. Each plot had 79% heather, 18% grass & 3% network of paths Grass divided into c118 patches per plot, 1sqm-690sq.m in area, most <10sq.m
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 36 yearling female Scottish Blackfaces from a single flock used. Allocated at random into 6x6 groups, individually marked for distance id. Each group allocated to one plot. 'Since groups of sheep stayed in the same plots throughout the experiment, effects of plot and group were necessarily confounded'
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: Each plot had 250 scan samples of positions and behaviours of individual sheep carried out during a 2-week observation period – 25 scans between 7:30 – 21:30/day for 11 days (extra day to account for low cloud obscuring plots) 'Behaviour recorded as grazing, lying, standing, walking with head up or drinking.'
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: Very short experimental period
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐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: Scotland, with Scottish Blackface sheep is not necessarily exactly the same as English uplands with local sheep breeds
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: Scanned observations by telescope from approx.500m marked directly onto vegetation maps and transferred to GIS maps of the plots
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: 'To allow for possible problems due to over-dispersion, patch ID was also included as a random effect in the model since distinguishing features other than the area of a patch could have contributed to the variation in the data'
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: Each plot had 250 scan samples of positions and behaviours of individual sheep carried out during a 2-week observation period – 25 scans between 7:30 – 21:30/day for 11 days (extra day to account for low cloud obscuring plots)
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: sufficient details given but not necessarily 100% transferable from Scottish Blackface to English breeds of sheep

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	Smith R S, Charman D, Rushton S P, Sanderson R A, Simkin J M, Shiel R S
	Year	2003
	Aim of study	Vegetation change in an ombrotrophic mire in northern England after excluding sheep.
	Study design	2
	Quality score	=QA 5.1 The questions asked in the paper have not been answered, while answers appear to have been given to unasked questions. +
	External validity	=QA 5.2 Detrended Correspondence Analysis was used to provide a framework for comparison of the Butterburn Flow data with the regional variation in upland grazed and ungrazed mires. Changes in vegetation on mires surrounded by forestry plantations may be only partly contributable to the loss of grazing. Other possible factors are the hydrological impacts of the plantations and long term climate change.
Population and setting	Source population	Border mires, adjacent to Kielder forest.
	Eligible population	Large areas of open moorland with accessible vegetation ranging from intensively utilised Festuca/Agrostis grassland to less desirable blanket and raised mire.

	Inclusion and exclusion criteria	
	Setting	Butterburn Flow is the largest of the Border mires, adjacent to Kielder forest and too wet to afforest.
Methods of allocation to intervention/control	Methods of allocation	10x 20mx20m plots randomly located in a peripheral and a central zone on the mire.
	Intervention description	One member of each plot pair fenced in 1988 to exclude sheep. Vegetation in each plot sampled by dividing in half east/west and positioning 5x 1sq.m quadrats in each using random grid coordinates. They were marked and surveyed in Aug/Sept 1988 &1992 and July 2002 (precipitated by loss of sheep in 2001 foot and mouth
	Control/comparison description	
	Sample sizes	10x2x5 = 100 quadrats over 10x40m sq plots
	Baseline comparisons	
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Was the species composition of the vegetation on Butterburn Flow influenced by very low level sheep grazing so that moorland species were kept in check thereby increasing the area dominated by some of the ombrotrophic mire species?
	Secondary outcome measures	If there was a vegetation change, was it related to the species composition of the vegetation at the start of the trial, how was it distributed around the mire and how did it relate to the vegetation of the ungrazed mires within Kielder Forest?
	Follow-up periods	1988, 1992, 2002
	Methods of analysis	Detrended Correspondence Analysis

Results		12 spp occurred in more than 59% of the quadrats in all survey years. The main trend is from species typical of very wet ombrotrophic mires to those more associated with dry moorland. While many plots remained relatively static over the 14 years, 2 ungrazed plots shifted towards the dry end of the axis. This may have been as a response to low summer rainfall prior to 1992. 'The main conclusion to be drawn is that significant vegetation change only occurred in a limited part of the mire edges following cessation of grazing.' 'The restriction of such changes to the periphery may reflect the distribution of sheep on the mire, with the best most accessible grazing being at the edge' 'The lack of change following cessation of grazing over the wetter areas suggests that the current precautionary management of re-wetting mires by blocking ditches and natural stream headwaters is a valid strategy'
Notes	Limitations identified by author	
	Limitations identified by review team	The questions asked in the paper have not been answered, while answers appear to have been given to unasked questions.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	British Ecological Society, English Nature and the Forestry Commission

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	Vegetation change in an ombrotrophic mire in northern England after excluding sheep. Smith R S, Charman D, Rushton S P, Sanderson R A, Simkin J M, Shiel R S Applied Vegetation Science 6 pp261-270. 2003
Study Design Category	2
Assessed by & when	Alison Hiles 19/2/2013

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: Large areas of open moorland with accessible vegetation ranging from intensively utilised Festuca/Agrostis grassland to less desirable blanket and raised mire. Butterburn Flow is the largest of the Border mires, adjacent to Kielder forest and too wet to afforest.
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:10x 20mx20m plots randomly located in a peripheral and a central zone on the mire
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: One member of each plot pair fenced in 1988 to exclude sheep. Vegetation in each plot sampled by dividing in half east/west and positioning 5x 1sq.m quadrats in each using random grid coordinates. They were marked and surveyed in Aug/Sept 1988 &1992 and July 2002 (precipitated by loss of sheep in 2001 foot and mouth)

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: 10x 20mx20m plots randomly located in a peripheral and a central zone on the mire
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: One member of each plot pair fenced in 1988 to exclude sheep. Vegetation in each plot sampled by dividing in half east/west and positioning 5x 1sq.m quadrats in each using random grid coordinates. They were marked and surveyed in Aug/Sept 1988 &1992 and July 2002 (precipitated by loss of sheep in 2001 foot and mouth)
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: Surveys completed 1988, 1992 & 2002
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 species frequencies in randomised quadrats over 43 species and 100 quadrats measured 1988, 1992 & 2002
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: all measurements done at the same time
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: carried out at 6 year and 10 year intervals

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-treatment 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: The questions asked in the paper have not been answered, while answers appear to have been given to unasked questions.
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: Detrended Correspondence Analysis was used to provide a framework for comparison of the Butterburn Flow data with the regional variation in upland grazed and ungrazed mires. Changes in vegetation on mires surrounded by forestry plantations may be only partly contributable to the loss of grazing. Other possible factors are the hydrological impacts of the plantations and long term climate change.

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Clay	2009	2	UK (N England)	Upland	1-	Partially randomised replicated plock	Grazing is background level, unrandomised
Worrall	2007		UK (N England)	Upland	1-	Partially randomised replicated plock	three burning treatments and grazed and ungrazed, factorially. Grazing is background level for the unit
Worrall	2008		UK (N England)	Upland	1-	Partially randomised replicated plock	three burning treatments and grazed and ungrazed, factorially. Grazing is background level for the unit
Dennis	2008	3	UK (C Scotland)	Upland	1+	replicated randomised blocks	Six replicates of four grazing treatments. Part of Glen Finglas GRUB study
Evans	2005	2	UK (C Scotland)	Upland	1+	replicated randomized block study	Part of GRUB - six replicates of four treatments. Study took place infirst two years of treatments.
Evans	2006a	3	UK (C Scotland)	Upland	1+	replicated randomized block study	Part of GRUB - six replicates of four treatments. Study took place infirst three years of treatments. Updated by Pakeman (Pers com) 2012
Evans	2006b	3	UK (C Scotland)	Upland	1+	replicated randomized block study	Part of GRUB - six replicates of four treatments. Study took place infirst three years of treatments.
Garnett	2000		UK (N England)	Upland	1+	Partially andomised control plots	10-yearand unburned grazed and ungrazed at Hard Hill. Grazing treatment not randomised.
Hetherington	2000	2	UK (Wales)	Upland	1+	partially randomised replicated block	Two areas with strategically placed blocks in three areas, two normal practice
Hulme	1999	7	UK (Scotland)	Upland	1+	Randomised block	Three grazing treatments based on target sward heights.
Pakeman	2003	5	UK (NE	Upland	1+	replicated block	Two replicates per treatment
Vandenberghe	2009	5	UK (C Scotland)	Upland	1+	Randomized replicated block study	Two treatments used, mixed grazing and high sheep from three sites of two replicates at each site
Ward	2007	2 (treatments 50 yrs)	UK (N England)	Upland	1+	replicated randomized block study	Moor House but only relicates of 10-yr burn rotation and long-term ungrazed used in this study
Littlewood	2008	5	UK (NE	Upland	1++	replicated randomised	Four treatments including ungrazed control with 6 replicates.
Littlewood	2012	5	UK (Scotland)	Upland	1++	replicated randomised blocks	Six replicates of four grazing treatments. Part of Glen Finglas GRUB study
Pakeman & Nolan	2009	5-6 years	UK (Scotland, N England)	Upland	1++	Meta-analysis of a number of replicated control trials	based on previous studies, but similar in methods, exposure etc.
Poulton	2011	13	UK (England)	Upland	1++	Meta-analysis of a large number of field surveys	Overgrazing heather condition surveys of 141sites, some with repeat visits
Anderson &Radford	1994	8	UK (English midlands)	Upland	2-	Survey/ correlative study	Area wide grazing changes monitired on a series of transects in part of the area
Baines	1996	3	UK (N England, S and Cental Scotland)	Upland	2-	Paired moorland blocks	four combinations of light or heavy grazing with keepering or none, over 5 areas
Britton	2005	1	UK (Wales)	Upland	2-	Plot correlative study	Vegetatrion compostion and soil and plant chemistry in a number of plots. Grazing pressure assesed by dung counts
Clarke	1995a	1	Scotland)	Upland	2-	Replicated block	Unrandomised. Two replicates of each patch type treatment.
Clarke	1995b	2	UK (NE	Upland	2-	Replicated block	Unrandomised. Two replicates of each patch type treatment.
Critchley	2008	4	UK (N England)	Upland	2-	Unreplicated paddocks	Four grazing treatments, low and high sheep, plus and minus cattle. Based on ESA rates.
Douglas	2008	1	UK (C Scotland)	Upland	2-	replicated randomized block study	Part of GRUB - six replicates of four treatments. Only the intensive grazing treatment used. Study onlyone summer, in year following experiment set up
Ferriera	2005	1	Spain	Upland	2-	Unreplicated single	Comparing goat and sheep grazing
Fisher	1994	3	UK (Scotland)	Upland	2-	Unrepicated paddock	sheep, goat and mixed graazing treatments.

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Fryday	2001	long-term	UK (Scotland, N England)	Upland	2-	Exclosures and comparisons	Exclosed for up to 40+ years
Gordon	2001	5	UK (Wales)	Upland	2-	unreplicated paddock	grazing treatments are continuation of ESA rate study
Grant	1968	5	UK (S & NE Scotland)U	Upland	2-	Survey approach with comparison exclosures	Thirty burned sites open to grazing with comparison ungrazed exclosures
Grant	1996b	5	UK (Scotland)	Upland	2-	Unreplicated block	one cattle, two sheep treatments. Separate study with three goat treatments and sheep control
Hartley	1997	3	UK, (NE	Upland	2-	Replicated block	grazing only presence/ absenbce
Hester & Bailey	1998	6	UK (NE	Upland	2-	Replicated block (non-	Two replicates of three treatmets, deer, sheep and mixed
Hester et al	1999	5	UK (NE	Upland	2-	Replicated block (non-	Two replicates of three treatmets, deer, sheep and mixed
Jewel	2005	6	Switzerland	Alpine	2-	unreplicated paddock	Grazing re-introduction - cattle
Milligan	2004	5	UK (N England)	Upland	2-	replicated randomised	Two replicates per treatment
Pearce-Higgins	2002		UK (N England, S Scotland)	Upland	2-	Stratified random sample	bird observations and habitat variables in 76 1km sqaures
Rawes	1983	14	UK (N England)	Upland	2-	unreplicated block	Two study sites, with grazed comparisons.
Ross	2000	3	UK (N England)	Upland	2-	Unreplicated controlled	Two ages of heather burned, with two grazing levels imposed on each
Rushton	1996	5	UK (Wales, N England)	Upland	2-	Unreplicated plot, field and farm scale	Part of ADAS Hills and Uplands study. Plot and field experiments at ESA and ESA-30% stocking. Testing of modelled veg change rate
Van der Wal	2003	4	UK (Scotland)	Upland	2-	replicated block, correlative	Various elements, experimental N addition, separate grazing exclusion and grazing effects in correlative study
Welch	1966	1	UK (N England)	Upland	2-	Comparative grazing unit study	Sheep counts over 14 months on three moorland plots grazed by farm hill flocks
Welch	1984a	7	UK (NE Scotland)	Upland	2-	Correlative study	32 sites with study plots of 0.4-2ha, identifed as range restricted ir unrestricted, withaccess to improved grass and no grass
Welch	1984b	7	UK (NE Scotland)	Upland	2-	Correlative study	32 sites with study plots of 0.4-2ha, identifed as range restricted ir unrestricted, withaccess to improved grass and no grass
Welch	1985	7	UK (NE Scotland)	Upland	2-	Correlative study	32 sites with study plots of 0.4-2ha, identifed as range restricted ir unrestricted, withaccess to improved grass and no grass
Welch & rawes	1964	7	UK (N England)	Upland	2-	Exclosures and comparisons	Three exlosures on high level habitat, grazed comparison areas with no baseline
Adamson &	2003	30+	UK (N England)	Upland	2+	Comparative plot study	Number of paired grazed and ungrazed plots
Albon	2007	7	UK (Scotland)	Upland	2+	Large-scale survey,	Different areas surveyed in eacjh year - one-off surveys
Amar	2011	9	UK (Orkney)	Moorland	2+	Correlative study	Moorland wide survey. Vegetation sampling in 18 1km squares
Bargett	2001	1 (varying length treatments)	UK (N England, Wales)	Upland	2+	Partially replicated and randomized block	Six grazing pressures identfied in three locations. A degree of replication within treatment sites
Calladine	2002	1	UK (N England)	Upland	2+	Paired grazing units	Treatment sites are agri- env reduced grazing sites vs non ag.
Cole	2010	3	UK (Scotland)	Upland	2+	Paddock comparison	2 paddocks, year-round sheep and summer only
Common	1988	5	UK (S	Upland	2+	Replicated block, no	Two cattle grazing treatments
DeGabriel	2011	3	UK (Scotland)	Upland	2+	Paired grazing units	Eight pairs of sheep/ deer paired with deer only (sheep removed up to 49 years to 5 years previously
Deleglise, BAE	2011	20	France (W alps)	Alpine	2+	Paired plots, 9	Uses pre-existing lon-term ungrazed plots
Deleglise, JVS	2011	20	France (W alps)	Alpine	2+	Paired plots, 9	Uses pre-existing lon-term ungrazed plots
Evans	1977	2+	UK (England, midlands)	Upland	2+	Single catcment study	Survey od previously initiated erosion. Recovery follwed up 6 years later
Fraser	2009	2	UK (Wales)	Upland	2+	replicated randomized block study?	Two sites, low and high Cv. Four sub plots with two breeds of sheep and 2 breeds cattle rotated around. Animals are the replicates, not plots, as diet study. One site in each year.

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Fraser	2011	8	UK (Wales)	Upland	2+	replicated block experiment	Previously ungrazed site. Treatments compare sheep and cattle summer grazing
Gardner	2002	10	UK (Wales)	Upland	2+	unreplicated plot study	Three sets of plots in two experiments lookng at original ESA rates, and seasonal effects.
Grant	1985b	3	UK (S & C Scotland)	Upland	2+	unreplicated plot study	three 3ha paddocks, with adjacent run-in plots. Different grass dominants. Mixed cattle and sheep
Grant	1987	4	UK (N Scotland)	Upland	2+	unreplicated plot study	A blanket bog and a heath site, one plot each. Animals grazed for different periods and rotated round plots. Part of larget study including Grant 1985b
Grant	1996a	6	UK (S	Upland	2+	Partially replicated block	Two cattle grazing treatments. Main grazing treatment unreplicated
Grant	1985	11	UK (W Scotland)	Upland	2+	unreplicated plot study	three sites of three plots with interactions of grazing level, seasonal pattern and time since burning.
Grant& Hunter	1968	11	UK (NE Scotland)	Upland	2+	unreplicated plot study	six plots with one cobination of H, L, winter, summer or yr round grazing, and 4 ages burning
Hartley	2005	6	UK (NE	Upland	2+	Replicated block	Presence/ absence grazing study. Sites per Hartley 1997
Hodgeson	1991	4	UK (S & C	Upland	2+	unreplicated plot study	Three grassland and two dwarf shrub plots, 3ha each
Hope	1996	up to 25	UK (Scotland)	Upland	2+	Comparative grazing unit study	Eleven paired sites, one of each pair having sheep reductionsfor up to 25 years
Hunter & Milner	1963	4	UK (S	Upland	2+	observational study	One grazing unit
Keiller	1995	1	UK (Wales)	Upland	2+	replicated block	based around pre-existing long term exclosures. One site not replicated
Kirkham & Milne	2000	3	UK (England & Wales	Upland	2+	replicated block experiment	Six sites and differnet vegetation types
Lawrence &	1998	1.5	UK (S	Upland	2+	observational study	One grazing unit, sheepobservational study
Littlewood	2006a	1	UK (England & Scotland)	Upland	2+	Correlative grazing unit study	4 grazing exclusion and 4 re-seeding restoration sites, inverts
Littlewood	2006b	2	UK (England & Scotland)	Upland	2+	Correlative grazing unit study	4 grazing exclusion and 4 re-seeding restoration sites, plant assemblage
Marrs	2004	5	UK (N England)	Upland	2+	Replicated block	Two grazing rates and absence, with interactions with burning and herbicide treatment
Martin	2010	8	UK (N England)	Upland	2+	Monitoring study	one site measired four times over 8 yrs
Medina-Roldan	2011	1	UK (N England)	Upland	2+	Comparative	measurements after 8 years of stock exclusion
Miller	1999	10	UK (S Scotland)	Upland	2+	Comparative plot study	Subjectivey placed to cover G nivalis clusters. Paired grazed and ungrazed summer
Miller	2010	10	UK (S Scotland)	Upland	2+	Comparative plot study	Subjectivey placed to cover G nivalis clusters. Paired grazed and ungrazed summer
Milne	2002	3	UK (England & Wales	Upland	2+	replicated block study	Six sites and differnet vegetation types
Oom	2008	3	UK (NE	Upland	2+	Stratified random block	Three sheep grazing treatments applied to each of two plots
Oom	2010	4	UK (NE	Upland	2+	replicated block	Two replicates per treatment, three sheep stocking rates
Palmer	2003	1	UK (NE	Upland	2+	Correlative study	Six sites based on estimated deer density, 2 x 3 levels
Pearce-Higgins	2006	2	UK (S Scotland, N England)	Upland	2+	Survey/ correlative study	Large sample of 85 plots over 10 upland areas. Correlative modelling approach
Rawes & Hobbs	1979	18+	UK (N England)	Upland	2+	Unreplicated plot	thirteen exclosures of varying age, comparisons with adjacent grazed areas.
Sibbald	2008	1	UK(NE	Upland	2+	replicated randomised	Groups of ewes allocated at random to eac of six blocks
Smith	2003	14	UK (N England)	Upland	2+	random paired plots	Five pairs, fenced and unfenced
Uff	2011	12	UK (W	Upland	2+	Corrlelative monitoring	Repeat heather condition monitoring

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Welch	1998	6	UK (England, midlands)	Upland	2+	Replicated block, non-randomised	bilberry and heather-bilberry sites
Welch	2006	10	UK (N Scotland)	Upland	2+	Survey/ correlative study	Response of heather over a period of deer grazing reduction in two Scottish glens.
Welch	2005	12	UK (NE Scotland)	Upland	2+	Transect correlative study	Effects on grazing and veg change as result of snow fence. See Van der Wal, 2003
Williams	2011	2	Eire	Upland	2+	One grazing unit study,	Sheep were the replicates, but they will interact
Cooper	1997	1	UK (N Ireland)	Upland	2++	Corrleative sample/ classification study	Stratified randon vegetation sample of upland Land Cover classes, and collection of data of environmetal and management variables
Critchley	in press	8	UK (Wales)	Upland	2++	randomised replicated	Factorial -disturbance and seeding treatments. Sites at Pwllpeiran
Dennis	1997	2	UK (S Scotland)	Upland	2++	replicated non-randomised plots	Two replicates of 2x target sward heights, each with sheep only and sheep + cattle treatment. Treatments in place for 5 years Stocking rates varied continually to maintain sward heights.
Dennis	2002	2	UK (S Scotland)	Upland	2++	replicated non-randomised plots	Two replicates of 2x target sward heights, each with sheep only and sheep + cattle treatment. Treatments in place for 5 years Stocking rates varied continually to maintain sward heights.
Dennis	2001	2	UK (S Scotland)	Upland	2++	replicated non-randomised plots	Two replicates of 2x target sward heights, each with sheep only and sheep + cattle treatment. Treatments in place for 5 years Stocking rates varied continually to maintain sward heights.
Hill	1992	13-25	UK (Wales)	Upland	2++	Replicated block	9 sites, winter grazing treatments in place for cariable amounts of time. Grazing exlusion is main focus.
Hulme	2002	6	UK (N England)	Upland	2++	replicated blocks	Two replicates of four treatments and control
Littlewood	2006c	1	UK (England & Scotland)	Upland	2++	Multi-site correlative grazing unit study	At each of six sites, six sample areas established in grass and 6 in heath
Mitchell	2008	5	UK (N England, Wales)	Upland	2++	randomised replicated block	Factorial -disturbance and seeding treatments. Sites at Pwllpeiran and Redesdale
Jenkins & watson	2001	Surveys over 41 year period	UK (NE Scotland)	Upland	3-	Observational study	Large moorland block surveyed over two periods in late 1950s and late 80's/ 90s and compared
Anderson & Yalden	1981	40	UK (English midlands)	Upland	3+	Mapping survey	Limited analysis if veg change from maps and co-incidental sheep data at Parish scale.
Evans	2005	30	UK (England, midlands)	Upland	3+	Single catcment study	Update of above
Johnston	2012	10+	UK (N England)	Upland	3+	Case study/ observations	Based on experience and observations or range of AE schemes. Some CSM data
Roberts	2002,2003,	10	UK (N England)	Upland	3+	Observational study	Based on thourough and fairly systematic botanical recording visits
Webb	2012	8	UK (N England)	Upland	3+	Case study/ observations	Observations on a singel agri-environment agreement

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Amar	2011	9	UK (Orkney)	Moorland	2+	Correlative study	Moorland wide survey. Vegetation sampling in 18 1km squares
Anderson & Yalden	1981	40	UK (English midlands)	Upland	3+	Mapping survey	Limited analysis if veg change from maps and co-incidental sheep data at Parish
Anderson &Radford	1994	8	UK (English midlands)	Upland	2-	Survey/ correlative study	Area wide grazing changes monitired on a series of transects in part of the area
Baines	1996	3	UK (N England, S and Central Scotland)	Upland	2-	Paired moorland blocks	four combinations of light or heavy grazing with keepering or none, over 5 areas
Britton	2005	1	UK (Wales)	Upland	2+	Plot correlative study	Vegetatrion compostion and soil and plant chemistry in a number of plots. Grazing pressure assesed by dung counts
Calladine	2002	1	UK (N England)	Upland	2+	Paired grazing units	Treatment sites are agri- env reduced grazing sites vs non ag.
Clarke	1995	1	UK (NE Scotland)	Upland	2+	Replicated block	Unrandomised. Two replicates of each patch type treatment.
Cole	2010	3	UK (Scotland)	Upland	2++	Paddock comparison survey	2 paddocks, year-round sheep and summer only
Common	1988	5	UK (S Scotland)	Upland	2+	Replicated block, no control	Two cattle grazing treatments
Cooper	1997	1	UK (N Ireland)	Upland	2++	Corrleative sample/ classification study	Stratified randon vegetation sample of upland Land Cover classes, and collection of data of environmetal and management
Critchley	2008	4	UK (N England)	Upland	2-	Unreplicated paddocks	Four grazing treatments, low and high sheep, plus and minus cattle. Based on
DeGabriel	2011	3	UK (Scotland)	Upland	2-	Paired grazing units	Eight pairs of sheep/ deer paired with deer only (sheep removed up to 49 years to 5 years previously
Deleglise, BAE	2011	20	France (W alps)	Alpine	2+	Paired plots	Uses pre-existing lon-term ungrazed plots

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Dennis	2008	3	UK (Scotland)	Upland	1+	replicated randomised blocks	Six replicates of four grazing treatments. Part of Glen Finglas GRUB study
Evans	1977	2+	UK (England, midlands)	Upland	2+	Single catcment study	Survey od previously initiated erosion. Recovery follwed up 6 years later
Gordon	2001	5	UK (Wales)	Upland	2-	unreplicated paddock study	grazing treatments are continuation of ESA rate study
Grant	1996	6	UK (Scotland)	Upland	2+	Partially replicated block	
Grant	1968	5	UK (S & NE Scotland)U	Upland	2-	Survey approach with comparison exclosures	Thirty burned sites open to grazing with comparison ungrazed exclosures
Grant	1985	11	UK (W Scotland)	Upland	2+	unreplicated plot study	three sites of three plots with interactions of grazing level, seasonal pattern and time since burning.
Hartley	1997	3	UK, (NE Scotland)	Upland	2-	Replicated block	grazing only presence/ absenbce
Hartley	2005	6	UK (NE Scotland)	Upland	2+	Replicated block	Presence/ absence grazing study. Sites per Hartley 1997
Hulme	1999	7	UK (Scotland)	Upland	1+	Randomised block, controlled	Three grazing treatments based on target sward heights.
Littlewood	2006	1	UK (England & Scotland)	Upland	2+	Correlative grazing unit study	4 grazing exclusion and 4 re-seeding restoration sites
Littlewood	2012	5	UK (Scotland)	Upland	1++	replicated randomised blocks	Six replicates of four grazing treatments. Part of Glen Finglas GRUB study
Littlewood	2008	5	UK (NE Scotland)	Upland	1++	replicated randomised control blocks	Four treatments including ungrazed control with 6 replicates.
Marrs	2004	5	UK (N England)	Upland	2+	Replicated block	Two grazing rates and absence, with interactions with burning and herbicide
Milligan	2004	5	UK (N England)	Upland	1+	replicated randomised blocks	Two replicates per treatment

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Oom	2010	4	UK (NE Scotland)	Upland	2+	replicated block	Two replicates per treatment, three sheep stocking rates
Pearce-Higgins	2002		UK (N England, S Scotland)	Upland	2-	Stratified random sample	bird observations and habitat variables in 76 1km sqaures
Pearce-Higgins	2006	2	UK (S Scotland, N England)	Upland	2+	Survey/ correlative study	Large sample of 85 plots over 10 upland areas. Correlative modelling approach
Pollock	2007	1	UK (Scotland, N England)		4++	Analysis of expert opinion	Analysis of 9 expert responses to questions on animal grazing preferences
Ross	2000	3	UK (N England)	Upland	2-	Unreplicated controlled plot study	Two ages of heather burned, with two grazing levels imposed on each
Rushton	1996	5	UK (Wales, N England)	Upland	2-	Unreplicated plot, field and farm scale	Part of ADAS Hills and Uplands study. Plot and field experiments at ESA and ESA-30% stocking. Testing of modelled veg change
Van der Wal	2003	4	UK (Scotland)	Upland	2-	replicated block, correlative	Various elements, experimental N addition, separate grazing exclusion and grazing effects in correlative study
Vandenberghe	2009	5	UK (C Scotland)	Upland	1+	Randomized replicated block study	Two treatments used, mixed grazing and high sheep from three sites of two replicates at each site
Welch	1998	6	UK (England, midlands)	Upland	2+	Replicated block, non-randomised	bilberry and heather-bilberry sites
Welch	1966	1	UK (N England)	Upland	2-	Comparative grazing unit study	Sheep counts over 14 months on three moorland plots grazed by farm hill flocks
Welch	2005	12	UK (NE Scotland)	Upland	2+	Transect correlative study	Effects on grazing and veg change as result of snow fence. See Van der Wal, 2003
Welch	2006	10	UK (N Scotland)	Upland	2+	Survey/ correlative study	Response of heather over a period of deer grazing reduction in two Scottish glens.
Welch & rawes	1964	7	UK (N England)	Upland	2-	Exclosures and comparisons	Three exlosures on high level habitat, grazed comparison areas with no baseline

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Douglas	2008	1	UK (C Scotland)	Upland	2-	replicated randomized block study	Part of GRUB - six replicates of four treatments. Only the intensive grazing treatment used. Study onlyone summer, in year following experiment set up
Evans	2005	2	UK (C Scotland)	Upland	1+	replicated randomized block study	Part of GRUB - six replicates of four treatments. Study took place infirst two years of treatments.
Evans	2006a	3	UK (C Scotland)	Upland	1-/+	replicated randomized block study	Part of GRUB - six replicates of four treatments. Study took place infirst three years of treatments. Updated by Pakeman
Evans	2006b	3	UK (C Scotland)	Upland	1+	replicated randomized block study	Part of GRUB - six replicates of four treatments. Study took place infirst three years of treatments.

Author	Publication date	Length of study (in years)	Country	Locality	Score	Study design	Notes
Hulme	2002	6	UK (N England)	Upland	2++	replicated blocks	Two replicates of four treatments and control
Pakeman	2003	5	UK (NE Scotland)	Upland	1+	replicated block	Two replicates per treatment
Pakeman & Nolan	2009	5-6 years	UK (Scotland, N England)	Upland	1++	Meta-analysis of a number of replicated control trials	based on previous studies, but similar in methods, exposure etc.
Poulton	2011	13	UK (England)	Upland	1++	Meta-analysis of a large number of field surveys	Overgrazing heather condition surveys of 141sites, some with repeat visits

Evidence Table

Name of Evidence Review:		Natural England Uplands Evidence Review
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		d) Over what timescales can grazing-related change in plant structure and diversity be observed or expected?
Study details	Authors	**1997: Sustainable Moorland Management Project - Progress Report. W. Jenkins**. **1998: Long Mynd Moorland Monitoring and Management Scheme. W. Jenkins** **1999 Moorland Management Project Report. W. Jenkins and P. Anderson**. **2000: Moorland Management Project Report. W. Jenkins and P. Anderson**. **2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011: Moorland Management Project Reports. Caroline Uff**.
	Year	1997 – 2012 (see above),
	Aim of study	to monitor the effects of grazing on The Long Mynd with a view to establishing a sustainable stocking level
	Study design	Correlation (correlating heather condition, grazing index and heather consumption with stocking levels)
	Quality score	2+
	External validity	EV+
Population and	Source population	Long Mynd Upland Heath, Upland flushes (M6, M10, M29), Acid grassland (U1)

setting	Eligible population	Sheep counts carried out across 2086ha. 35 plots sampled for heather consumption.
	Inclusion and exclusion criteria	Not known.
	Setting	Long Mynd upland heath – NVC communities not specified. Upland flushes (M6, M10, M29), Acid grassland (U1)
Methods of allocation to intervention/control	Methods of allocation	The 35 plots where heather consumption measurements are taken are in similar areas of the hill every year. Heather consumption measurements made every year bar 2001. Sheep counts are also made following a similar route every time. Sheep counts not made every year.
	Intervention description	Sheep numbers set using cross-compliance (1997, 1998, 1999 & 2000 reports) and ESA prescriptions (2001 report and all reports thereafter). 3.5ewes/ha in summer and 2.5ewes/ha in winter for reports at the start of the sequence 1997 going down to 2.5/2.0 for 1999 report going down to 1.5/0.75 for 2002 report and all those thereafter.
	Control/comparison description	Comparing changes in heather consumption, heather grazing index and heather condition with changes in sheep numbers.
	Sample sizes	Sheep counts carried out across 2086ha. 35 plots sampled for heather consumption
	Baseline comparisons	Baseline is from the 1997 report.
	Study sufficiently powered	I don't know. The results (e.g. the percentage of grazed shoots and changes in sheep numbers) are presented in tables. (e.g. correlating changes in the percentage of grazed shoots and changes in sheep

		numbers) are examined mainly in a qualitative and subjective way. They are presented in tables and
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Heaths - % cover of dwarf shrubs; % flowered; % of shoots grazed; Year burnt; heather condition; Heather consumption as a percentage.. Flush – vascular plant species with DAFOR; dunging; moss cover; sphagnum cover; wetness; hoof-print damage
	Secondary outcome measures	Heather consumption as a percentage Levels of suppression on heather
	Follow-up periods	Once
	Methods of analysis	Subjective - visual comparisons of bar diagrams and graphs. Very little statistical analysis of the mountains of data, which could well benefit from some expert analysis to produce a paper of the unusually detailed and long-running analysis.
Results		The condition of existing heather improved but there were no significant increases in heather area. The 2011 report says 'the heather continues to do well. This reflects the appropriate grazing levels and negligible heather beetle impact. The detailed condition of heather and levels of grazing on the heath were reported on in the spring update (grazing impact lowest recorded and grazing damage to heather almost eradicated)'. This has been achieved with a stocking rate of 1.5 ewes/ha in summer and 0.75 in winter.
Notes	Limitations identified by author	Caroline Uff says 'There is a huge amount of data there that would make a great PhD or masters project, but I've never needed to analyse it beyond roughly making sure our management is delivering what we want. Once I'm happy with that the data just gets archived!'
	Limitations identified by	Very little statistical analysis of the mountains of data, which could well benefit from some expert analysis to produce a paper of the unusually detailed and long-running

	review team	analysis.
	Evidence gaps and/pr recommendations for further research	See above
	Sources of funding	National Trust, English Nature and Natural England (through HLS).

Summary

Uff (1997-2011) sought to monitor the effects of grazing on The Long Mynd with a view to establishing a sustainable stocking level. Stocking densities across the hill (against which heather condition and grazing index were correlated) were obtained by trusting the graziers to follow agreed stocking limits, backed up with periodic sheep counts across 85% of the area. Heather condition was assessed using a variety of measurements (including % of shoots grazed, % heather consumption, % flowering). Measurements were taken annually for 15 years on 35 plots spread evenly across the site's homogenous areas of heath.

Uff (1997-2011) The 2011 report says 'the heather continues to do well. This reflects the appropriate grazing levels and negligible heather beetle impact. In 2011 the average % of grazed shoots on heather was 11%. This has been achieved with a stocking rate of 1.5 ewes/ha in summer and 0.75 in winter.

Caroline Uff says 'There is a huge amount of data there that would make a great PhD or masters project, but I've never needed to analyse it beyond roughly making sure our management is delivering what we want. Once I'm happy with that the data just gets archived!'

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

Name of Evidence Review: Natural England Uplands Evidence Review

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	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 Citation	Uff, C (2011, 2009, 2008, 207, 2006, 2005, 2004, 2003, 2002, 2001) Moorland Management Project Report W. Jenkins and P. Anderson (2000, 1999) Moorland Management Project Report W. Jenkins (1998) Long Mynd Moorland Monitoring and Management Scheme W. Jenkins (1997) Sustainable moorland management project – progress report.
Study Design Category	2
Assessed by & when	Tom Holland 6th February 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.	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Yes – Long Mynd Upland Heath, Upland flushes (M6, M10, M29), Acid grassland (U1)
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Yes.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Yes. 35 transects and sheep counts across 85% 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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: I think the plots were selected to give an even spread of plots across areas of homogenous heather across a range of hefts.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Seemed reasonable to me though some of the assessment methods have fallen out of favour over the years. That is, English Nature's Heather Grazing Index and the estimates of heather consumption. Heather grazing index is still used in Common Standard Monitoring guidance.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: 1997 when stocking densities were still relatively high forms the baseline (measured in the same way as subsequent surveys).
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Year on year climatic variation difficult to take into account. N-deposition similarly difficult to take into account.
2.5 Is the setting applicable to the UK?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Yes

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: A lot of the measures are prone to differences in observer bias 7 error (e.g. % cover assessments, heather condition; heather consumption, DAFOR scores). However, some of the most useful measurements (such as heather shoots grazed and sheep numbers) should be less prone to observer 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?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: English Nature Grazing Indx survey dropped fairly early on but some of the most useful measurements (such as heather shoots grazed and sheep numbers) have been measured annually for 15 years.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Caroline Uff says 'There is a huge amount of data there that would make a great PhD or masters project, but I've never needed to analyse it beyond roughly making sure our management is delivering what we want. Once I'm happy with that the data just gets archived'.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Not sure.
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Some of the most useful measurements (such as heather shoots grazed and sheep numbers) have been measured annually for 15 years

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: The study has been going for 15 years with useful measurements (such as heather shoots grazed and sheep numbers) measured annually, but as Caroline Uff says 'There is a huge amount of data there that would make a great PhD or masters project, but I've never needed to analyse it beyond roughly making sure our management is delivering what we want'.

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ☐**NR** ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: No statistical analysis.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++ ☐+ ☐- ☐NR ☐NA	Comments: Probably does what it was designed to do but the analytical methods that have used are subjective - visual comparisons of bar diagrams and graphs. Very little statistical analysis of the mountains of data, which could well benefit from some expert analysis to produce a paper of the unusually detailed and long-running analysis.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐++ ☐+	Comments: No statistical analysis.

Were confidence intervals and or p-values for the effect estimates given or calculable?	☐- ☐NR ☐NA	
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ☐-	Comments: Some of the most useful measurements (such as heather shoots grazed and sheep numbers) are less prone to bias than some of the other less useful measurements (e.g. % cover scores of various variables).
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: Probably applicable to other areas of upland heath and flush in England.

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	Vandenberghe, C., Prior, G., Littlewood, N.A., Brooker, R. & Pakeman, R.
	Year	2009
	Aim of study	To investigate foraging site selection of meadow pipits in response to two different grazing regimes
	Study design	1
	Quality score	+
	External validity	++
Population and setting	Source population	Not much background information on moorland skylark populations for example typical breeding densities etc. More contextual information given – past studies.
	Eligible population	The vegetation of the wider site is given in terms of broad communities. It is stated that the site is typical of many parts of upland Scotland
	Inclusion and exclusion criteria	Selection not described. Site and selection may be better described elsewhere as it is part of the GRUB study (e.g. Dennis et al, 2008). However three sites each of two replicate blocks, so probably chosen to reflect the range of typical vegetation on the

		site.
	Setting	Glen Finglas, central Scotland. Site between 200 and 500m.
Methods of allocation to intervention/control	Methods of allocation	Each block consisted of four plots, with four grazing treatments randomly assigned in each block
	Intervention description	The four grazing treatments of high sheep (nine sheep per 3ha plot) and low sheep (three sheep per plot), mixed grazing (two sheep and twp cows) and lack of grazing are well described. **However in this study only high sheep and mixed grazing plots were used, to obtain sufficient nest replicates for observation.** Sheep remained on plot throughout year other than when taken in for normal farming operations and periods of severe weather. Cattle grazed on mixed plot in Sept and Oct.
	Control/comparison description	None as such. The low sheep treatment is said to reflect the previous farm grazing regime, but is not considered in this study, and high sheep typical of current commercial rates for the habitat/ situation.
	Sample sizes	Two sites used due to high predation rate at the third. 44 paired squares, one of each pair with a foraging point. Vegetation cover in each square, and height and density measurements on a nine-point grid in each square.
	Baseline comparisons	N/A.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and	Primary outcome measures	Vegetation height, density and composition; foraging distance; invertebrate abundance and biomass
	Secondary outcome measures	Shannon diversity index for vegetation

significance)	Follow-up periods	Treatments in place for 5 years when this study took place
	Methods of analysis	For most variables differences between grazing treatments tested using residual maximum likelihood estimation. Number of invertebrate groups between grazing treatment and square types (foraging vs random) tested using non-parametric Kruskal-Wallis test.
Results		Vegetation height and density were significantly higher in low sheep and cattle than in the high sheep grazed plots ($p<0.0001$). Under both grazing treatments height ($p<0.001$) and density ($p<0.01$) were significantly higher in random than foraged squares. Differences were similar in both treatments. Foraged squares were more diverse ($p<0.05$) with Molinia cover being lower. Other differences in species cover were small. There were no significant differences in invertebrate group composition between treatments. Total abundance was higher in the mixed grazed plots ($p<0.05$) but did not differ significantly between square types. The pattern in total biomass was similar. Whilst invertebrate groups differed significantly in abundance, total biomass and mass per individual, there were no significant differences within groups between forage and random squares. Total biomass increased significantly with height class, but was not significantly different between square types. Although not significant, the difference between square types in total invertebrate biomass tended to decrease with increased vegetation height. Meadow pipits tended to forage in areas with lower vegetation height and density and with a lower proportion of the dominant tussock-forming grass Molinia caerulea. They did not forage in areas with a total higher invertebrate biomass but the foraging sites in the preferred vegetation type tended to have higher invertebrate biomass than similar vegetation at random sites. Foraging distance was greater in the more heavily grazed plots. Food accessibility seems to become an even more important criterion under high grazing intensity, where prey abundance and size decrease. In this study a low intensity mixed grazing regime seemed to provide a more suitable combination of sward height

		and structural diversity and food supply for foraging meadow pipits than more intensive sheep-dominated grazing.
Notes	Limitations identified by author	Single sampling method captured only a portion of available invertebrate biomass. Leatherjacket density (a significant part of the diet) was low and variable so could not be tested significantly. May be a factor of time lag between observations and sampling.
	Limitations identified by review team	Limited comparison of grazing regimes as only two treatments with sufficient nesting success; one season of measurements so didn't take account of seasonal climatic effects.
	Evidence gaps and/pr recommendations for further research	To quantify provisioning rates, prey biomass and fledging rates between grazing treatments to provide further insight to the effects of grazing on condition of birds during the breeding season
	Sources of funding	Scottish Government Rural and Environmental Research and Analysis Directorate

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	Vandenberghe, C., Prior, G., Littlewood, N.A., Brooker, R. & Pakeman, R. (2009). Influence of livestock on meadow pipit foraging behaviour in upland grassland. Basic and Applied Ecology, 10, 662-670
Study Design Category	1
Assessed by & when	D Martin 2/01/13

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 much background information on moorland skylark populations for example typical breeding densities etc. More contextual information given– past 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 vegetation of the wider site is given in terms of broad communities. It is stated that the site is typical of many upland parts of Scotland.
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?0	☐+	Comments: Selection not described. Site and selection may be better described elsewhere as it is part of the GRUB study (e.g. Dennis et al, 2008). However three sites each of two replicate blocks, so probably chosen to reflect the range of typical vegetation on the site.

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	☐++	Comments: Each block consisted of four plots, with four grazing treatments randomly assigned in each block

likely/not likely?		
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: The four grazing treatments of high and low sheep, mixed grazing and lack of grazing are well described. However in this study only high sheep and mixed grazing plots were used, to obtain sufficient nest replicates for observation.
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: Foraging study carried out in 2007 in the fifth year of grazing treatments, likely to have been long enough to allow treatment effects on vegetation structure and associated invertebrate communities to develop.
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: Not reported but likely to have been low
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐NR	Comments: Unlikely, although there may have been some variation in grazing regimes through removal in severe weather
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Vegetation types are widespread in uplands and sheep grazing is normal upland land use so likely to be representative.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: The low sheep treatment is said to reflect the previous farm grazing regime, and high sheep typical of current commercial rates for the habitat/ situation.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures	☐++	Comments: Foraging behaviour based on observation of meadow pipit nests. Foraging points (where birds observed to fly from with food) paired with random points at same distance from nest. In quadrats around foraging and random points inverts (abundance and biomass) were sampled systematically for standard time using suction sampling. Vegetation height and

(e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?		density measured using drop discs and marked stick, and cover of species or groups estimated as proportions.
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 times between behavioural observations and field sampling varied, partly due to effects of weather.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: Only two grazing regimes compared due to low nest numbers in low sheep-grazed and ungrazed treatments.
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: Although no real comparison treatment as no ungrazed or low sheep treatment used
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments:

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐NR	Comments: Baseline conditions (at introduction of treatments) not reported here, but likely to be reported in other papers associated with studies at this site e.g. Dennis et al (2008) or Evans et al (2006)
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		Comments:

or calculable?	☐NR	
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: For most variables differences between grazing treatments tested using residual maximum likelihood estimation. Number of invertebrate groups between grazing treatment and square types (foraging vs random) tested using non-parametric Kruskal-Wallis test.
4.5 Was the precision of the intervention effects given or calculable? Were they meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++	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 and replicated design. Unfortunately not all treatments had sufficient nesting to use in this 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:

Dennis et al (2008) The effects of livestock grazing on of foliar arthropods associated with bird diet in upland grasslands in Scotland. Journal of Applied Ecology 45, 279-287

Evans et al (2006) Low intensity mixed livestock grazing improves the breeding abundance of a common insectivorous passerine. Biology letters, 2, 636-638.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

Name of Evidence Review: __Uplands_____

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

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Van der Wal, R., Pearce, I., Brooker, R., Scott, D., Welch, D. & Woodin, S. (2003) Interplay between nitrogen deposition and grazing causes habitat degradation. Ecology Letters, 6, 141-146
Study Design Category	2
Assessed by & when	D Martin 30/11/12

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: Scottish montane habitats
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: High level moss-sedge (*R lanuginosum – C bigelowii* 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: Site selection not described, but likely to be a good example of the habitat. Part of the study (sheep grazing and vegetation change) based on an existing long-term study on effects of snow fence on grazing patterns.

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: In main N addition experiment assignment of treatments was randomised. Ten replicates per treatment. Ten grazing exclusion cages set up – no indication of selection. Sheep habitat use measured through 15 sets of 5 dung plots ranged across summit. Likely to have been systematic to cover the area. Monitoring of sheep impact on *Racomitrium* and graminoid cover measured in fixed plots on 15 transects (at regular intervals?) perpendicular to a snow fenceline (also described in Welch, 2005). This is a correlative approach.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐++	Comments: Five treatments, including distilled water control. Two forms of N addition, at high and low rates. One application per season? Effects of sheep exclusion on moss growth measured in ten $1m^2$ cages. Ten shoots in each of four netlon cylinders. Sampling of moss shoots and vegetation sampling (pin-frame quadrats) well described.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: Treatment applied for four seasons. Vegetation impacts of sheep grazing measured on the transects over a 6 year period.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: No contamination reported
2.5 Were any other other intervention(s) received and, if so, were they similar in both groups? Did either group receive additional interventions (eg management not part of the experimental interventions, eg plots with unplanned burning)? Were groups treated equally?	☐+	Comments: Background N deposition – but would have been same across the treatment blocks
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: Habitat has limited extent in England, the study site is representative of summit plateau communities in NE Scotland. Main English examples are similar habitat, but much less extensive and may

		well be less diverse.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Control is distilled water (in reality also subject to the prevailing atmospheric N deposition), with background hill grazing level.

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: Replicated plots, vegetation measured objectively from point quadrats. Samples sizes generally large.
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: In accordance with the stated aims
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐-	Comments: No measure of relationship between N addition and grazing pressure in treatment plots. Paper assets that N addition will increase grazing pressure through favouring graminoids, but the sheep grazing effect element is correlative and does not explore causal relationships. Did sheep occupancy (dunging) increase on N plots?
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: Measures of moss N content and other related chemical properties, as well as growth effects.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐++	Comments: Yes, certainly in the main experimental treatments.
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: main experimental treatments in place for four years, long enough for effects to be detected. Correlative sheep grazing study measured over a six year period. Other aspects of the study (grazing exclusion, shading) carried out over shorter periods, but enough to allow significant effects to be detected.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐++	Comments: N addition treatment plots reported as not differing significantly in graminoid or *Racomitrium* cover at the start of the experiment. Background N and sheep grazing likely to have been broadly even, large number of replicates adjust for spatial variation. Sheep and vegetation change study is correlative approach so accounts for habitat variation.
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: Not really described. Largely ANOVA and correlative techniques to identify differences in N treatment effects, and grazing density effects from the different studies respectively. No interactions explored. The assertion that N addition is likely to increase grazing pressure is not really tested.
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, standard errors given in tables and graphs.
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?	☐-	Comments: Well designed and replicated in the most part, but background deposition does not seen to be taken into account, and limited exploration of relationship between N deposition and increased grazing.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)?	☐+	Comments: A number of findings relating to direct toxicity effects and change in cover balance are of direct relevance to the wider community.

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)?		

Welch, D., Scott, D. & Thompson, D.B.A. (2005) Changes in the composition of *Carex bigelowii – Racomitrium lanuginosum* moss heath on Glas Maol, Scotland, in response to sheep grazing and snow fencing. Biological Conservation 122, 621-631

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	Van der Wal, R., Pearce, I., Brooker, R., Scott, D., Welch, D. & Woodin, S.
	Year	2003
	Aim of study	To explore the interactions between N deposition and grazing in the degradation of montane moss-heath habitat.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	Scottish montane habitats
	Eligible population	High level moss-sedge (*R lanuginosum – C bigelowii* heath)
	Inclusion and exclusion criteria	Site selection not described, but likely to be a good example of the habitat. Part of the study (sheep grazing and vegetation change) based on an existing long-term study on effects of snow fence on grazing patterns.

	Setting	Glas Maol, Grampians, eastern Scotland. 1000m asl
Methods of allocation to intervention/control	Methods of allocation	In main N addition experiment assignment of treatments was randomised. Ten replicates per treatment. Ten grazing exclusion cages set up – no indication of selection. Sheep habitat use measured through 15 sets of 5 dung plots ranged across summit. Likely to have been systematic to cover the area. Monitoring of sheep impact on *Racomitrium* and graminoid cover measured in fixed plots on 15 transects (at regular intervals?) perpendicular to a snow fenceline (also described in Welch, 2005). This is a correlative approach.
	Intervention description	Five treatments, including distilled water control. Two forms of N addition, at high and low rates. One application per season? Effects of sheep exclusion on moss growth measured in ten $1m^2$ cages. Ten shoots in each of four netlon cylinders. Sampling of moss shoots and vegetation sampling (pin-frame quadrats) well described. Additional correlative study of sheep density (dung counts) and vegetation along transects from fence (shelter).
	Control/comparison description	In N addition study control is distilled water treatment. Studies of moss shoot growth from caged areas, with comparison as prevailing grazing conditions. On grazing transects there is a gradation from increased grazing near the fence to ambient grazing levels further away.
	Sample sizes	Ten replicates of each N addition treatment. Shoot growth from 10 shoots in four cylinders, in each of ten protected areas. Sheep grazing and graminoid abundance measured in 15 sets of 5 6m x 6m dung plots. Correlative studies of grazing and vegetation at five distances (dung plots and pin frames) along 15 transects.
	Baseline comparisons	N treatments were shown not to differ significantly in composition at start.
	Study sufficiently powered	No power analysis reported, but good sample sizes.

Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Graminoid and *Racomitrium* cover. Tissue N concentration, nitrate reductase activity and K leakage on shoot samples, shoot growth in absence of grazing, effects of shading. Relationship between sheep occupancy (dung) and vegetation change.
	Secondary outcome measures	
	Follow-up periods	Four years in N addition plots, grazing impacts measured over a 6 year period. Other measurements in one season.
	Methods of analysis	Not really described. Largely ANOVA and correlative techniques to identify differences in N treatment effects, and grazing density effects from the different studies respectively. No interactions explored. The assertion that N addition is likely to increase grazing pressure is not really tested
Results		N application resulted in significant loss of *Racomitrium* and increase in graminoid cover, with effects most marked at the high N treatments ($p<0.0001$). Direct toxicity was observed through increased tissue N and reduced N assimilation mechanisms, suggesting the moss was N saturated. K leakage was also significantly increased suggesting cell membrane damage. Shoot growth was significantly diminished. N addition has a direct fertilisation effect on grasses, resulting in reduced light availability for mosses. The effect of reduced light levels from shading was confirmed from greenhouse experiments. Increased grass cover also attracts sheep, with associated increases in trampling. Exclusion cages showed that *Racomitrium* growth was 40% lower in grazed plots. Along a marked gradient in sheep grazing density, generated by the sheltering effect of the snow fence, a marked decline in *Racomitrium* and corresponding increase in grass dominance was seen with increasing density, to almost no *Racomitrium* at densities estimated at 4 sheep per ha. The interaction between N deposition and grazing is a mult-step feedback loop, were toxicity to moss, graminoid fertilization, shading of moss and attraction of herbivores

		together lead to the replacement of moss-dominated vegetation by grasses and sedges.
Notes	Limitations identified by author	
	Limitations identified by review team	No indication of interplay with background N deposition, and how much N control is receiving. limited exploration of relationship between N deposition and increased grazing.
	Evidence gaps and/pr recommendations for further research	Review of critical loads of N, to take account of amplification effects of grazing.
	Sources of funding	NERC

Evidence Table

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

Study details	Authors	Ward, S. E., Bardgett, R. D., M^{C}Namara, N. P., Adamson, J. K. & Ostle, N. J. 2007. Long-term consequences of grazing and burning on northern peatland carbon dynamics. *Ecosystems,* 10**,** 1069-1083
	Year	2007
	Aim of study	To examine the long-term consequences of regular disturbance from controlled burning and grazing, on their own and in combination, on vegetation composition, C stocks, dissolved organic carbon (DOC) carbon dioxide (CO_2) and methane (CH_4) fluxes.
	Study design	1 Replicated, randomized design
	Quality score	+
	External validity	++
Population and setting	Source population	The extent of blanket peat. Not well described in terms of vegetation, but management practices on peat describes, and likely impacts on C.
	Eligible population	High level blanket bog at Moor House NNR. Likely to be fairly representative of upland blanket bog. Described as M19b in NVC terms.
	Inclusion and exclusion	The study site was an existing replicated burning and grazing exclusion experiment,

	criteria	established in 1954.
	Setting	Four blocks set over an area of 1km^2 at Moor House NNR in the North Pennines. Altitude of 590-630m, with sub-arctic oceanic climate. On peat 1-2m thick.
Methods of allocation to intervention/control	Methods of allocation	Not fully described, but reported elsewhere. It would appear that the three burning treatments were randomized within each of the four plots.
	Intervention description	Four combinations of treatments in factorial design. The long-term treatments included in this study are 10-year rotation burning and no burning, each with either grazing or no grazing. Four replicates of each. The 20-yr rotation plots not used in this experiment.
	Control/comparison description	The grazed, 10-yr burned plot is closest to typical management. However in the context of the aims of the study the ungrazed, unburned is the control treatment.
	Sample sizes	Four replicates. Veg composition from one small quadrat per plot, once per quarter. Carbon sampled in peat cores at 16 points per treatment. Other microbial and N measures from 5 cores at four periods. Gas fluxes at one point per plot monthly, and soil DOC at one point per plot monthly.
	Baseline comparisons	All areas burned at baseline. Expt commenced on recently burned vegetation. Similarity of plots prior to burning not known.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	C stocks, peat microbial activity and N availability, trace gas fluxes (CO_2, CH_4), DOC
	Secondary outcome measures	Vegetation composition, in terms of three functional groups

	Follow-up periods	Measurements over 18 month period, covering two spring/ summer periods
	Methods of analysis	Mixed model repeated measures ANOVA. Soil microbial activity, C stocks and vegetation ANOVA using GLM. Residuals checked for normality and transformed where necessary.
Results		Burning was shown to increase biomass of graminoids by 88% relative to unburned plots, but reduced the biomass if bryophytes (92%) and shrubs (51%). The shrub component was by far the greatest in terms of biomass in all treatments. The effect of grazing was similar, but smaller in magnitude. Shrub biomass was reduced by 18%, bryophytes by 47% and no effect on graminoids. Differences in C stocks were observed in aboveground vegetation and upper peat horizons only. The *F* and *H* layer (root zone, Litter layer) and above ground plant material contained around 60% less C in burned compared to unburned plots. Grazing reduced C in aboveground vegetation by 22%. There was no effect at 1m depth. There was no significant effect of grazing on soil microbial properties. Burning had the greatest effects on CO_2 fluxes. Grazing however increases rates of respiration and photosynthesis relative to ungrazed treatments, but to a lesser extent than burning. Grazed plots acted as a greater net sink for CO_2 than ungrazed plots over 10 of the 15 dates sampled. There were no significant interactions for grazing and burning on any of the measures. Seasonality accounted for more variation than land use treatment. Grazing significantly increased CH_4 effluxes at all sample dates compared with ungrazed, with the effects greater than for burning. The lowest fluxes occurred in the ungrazed, unburned plots. DOC was only affected by grazing, with greater concentrations at 10cm depth compared to ungrazed. Grazing has been shown to significantly affect above ground C storage, reducing it by 22% in light summer grazed plots compared with ungrazed. This can be attributed to

		the greater biomass of C-rich shrubs relative to graminoids. Grazing however increases rates of respiration and photosynthesis relative to ungrazed treatments, but to a lesser extent than burning. Grazed plots acted as a greater net sink for CO_2 than ungrazed plots over 10 of the 15 dates sampled. The results suggest that long term disturbance from burning and grazing increased ecosystem processes and gross CO_2 fluxes, and reduced net efflux. Grazing significantly increased CH_4 effluxes at all sample dates compared with ungrazed, although the reasons remain unclear, with the effects greater than for burning. The lowest fluxes occurred in the ungrazed, unburned plots. DOC was only affected by grazing, with greater concentrations at 10cm depth compared to ungrazed. The effect was small, and mechanisms unclear. The findings indicated that release of DOC was controlled by climate rather than land use. There was no detectable effect on soil microbial processes such as N mineralisation.
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	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	Ward, S. E., Bardgett, R. D., M^{C}Namara, N. P., Adamson, J. K. & Ostle, N. J. 2007. Long-term consequences of grazing and burning on northern peatland carbon dynamics. *Ecosystems,* 10**,** 1069-1083
Study Design Category	1
Assessed by & when	D Martin 17/01/13

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 in terms of vegetation, but management practices on peat describes, and likely impacts on C.
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: High level blanket bog at Moor House NNR. Likely to be fairly representative of upland blanket bog
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: Original allocation not described. Plots established in 1954. May have been bias in site selection.

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	☐++	Comments: Fully factorial grazing and burning experiment. Four replicates of three treatments, with the burning treatment appearing to have been randomized in the grazed and ungrazed plots.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Yes, burned on 10 or 20 yr rotation, and unburned since start. Half of each block has grazing excluded. This study used only the ten-year and unburned treatments, in grazed and ungrazed areas. Grazing level light, but not really quantified – was said to be 0.04 sheep ha^{-1} in summer.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐++	Comments: Yes. At time of study 9 years into 10 year burning cycle, so in 5^{th} cycle. Ungrazed treatment in place since 1954.
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:
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: The grazed, 10 year burned plots are most representative. However the grazing levels are lighter than typical

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: Community composition based on biomass sampling. Objective, but v small samples of 25cm^2, one from each plot every quarter. Peat cores to a depth of 1m taken for C measurement. Root and litter layer also sampled. Sixteen random cores per plot. Peat also sampled for microbial activity and N availability. Gas fluxes measured at monthly intervals from May 2003 and Sept 2004 using chambers. Soil DOC measured at different depths monthly, but from one point per plot.
3.2 Were all outcome measurements complete? Were outcome variables/measurements completed across all/most of the study population(s)/area(s) (that met the defined study outcome definitions)?	☐++	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments:
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:

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: This would have been reported in previous papers. All plots were burned when plots set up in 1954
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)?	☐NR	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?		
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: Mixed model repeated measures ANOVA. Soil microbial activity, C stocks and vegetation ANOVA using GLM. Residuals checked for normality and transformed where necessary.
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: Generally well designed and long-term. Some of the samples are limited (e.g. Veg biomas). Only two treatments compared, so longer term burning, as recommended on peatland, not included.
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐++	Comments:

Quality Assessment Checklist: 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	Webb, S (2012) The grazing impact on mountain vegetation, Glenridding Common, Helvellyn, Lake District. NE internal
Study Design Category	3
Assessed by & when	D Martin 30/12/12

Section 1: Theoretical approach		
1.1 **Is a qualitative approach appropriate?** For example: Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings? Could a quantitative approach better have addressed the research question?	☐ Appropriate	Comments: Summary of observations/ findings from long-term involvement in site management and associated management agreements. Some baseline monitoring but has not been repeated to date so these observations are best available information.
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: Simply to report observations on changes in vegetation following sheep reductions
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: in the context of lack of formal repeat monitoring data.

- 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?	☐ Defensible	Comments: Not a planned experiment/ project but observation based on sound knowledge of site. Repeat monitoring data not available.

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?	N/A	Comments:

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: Conservation Adviser responsible for setting up the grazing management agreement, with ten years experience of the site.
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:	☐ Unreliable	Comments: No data collection – observation.

-was data collected by more than one method? -is there justification for triangulation or for not triangulating? - do the methods investigate what they claim to?		

Section 5: Analyses		
5.1 Is the data analysis sufficiently rigorous? For example: -Is the procedure explicit? -how systematic is the analysis, is the procedure reliable? -is it clear how the themes and concepts were derived from the data?	☐ Not Rigorous	Comments: No analysis as such
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?	☐ Not Sure / Not Reported	Comments:
5.3 Is the analysis reliable? For example: -did more than one researcher theme and code data? -if so how were differences resolved? -were negative / discrepant results addressed?	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: No reason to doubt findings and it is highly likely there has been some observable effects, but findings not based on data collection
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	☐ Adequate	Comments:

interpretation and conclusions? -are the conclusions plausible and coherent? -have alternative explanations been explored and discounted? -does this enhance understanding of the research topic? -are the implications of the research clearly defined? -is there adequate discussion of the limitations encountered?		

Section 6: Ethics		
6.1 How clear and coherent is the reporting of ethics? 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: Observation and opinion based, but likely to give a reasonable indication of the most apparent/ easily observable changes in the site.

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	Webb, S
	Year	2012
	Aim of study	To report observations on observable changes in vegetation following sheep reductions
	Study design	3 – observations from a number of visits before and after stock reductions
	Quality score	+
	External validity	+
Population and setting	Source population	Upland grazing unit with a range of habitats including cliff ledge and flushes of varying base status
	Eligible population	As above
	Inclusion and exclusion criteria	Subject to sheep grazing reductions under management agreements

	Setting	Glenridding Common, part of Helvellyn and Fairfield SSSI, Lake District, Cumbria. Extends to summit of Helvellyn.
Methods of allocation to intervention/control	Methods of allocation	N/A
	Intervention description	Significant reduction in grazing from annual average of c0.14 LU/ha to 0.04 LU/ha including off-wintering.
	Control/comparison description	Previous higher stocking rate before reduction
	Sample sizes	N/A
	Baseline comparisons	N/A some baseline survey pre-2005
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation structure and condition
	Secondary outcome measures	
	Follow-up periods	Stock reductions in place for 8-9 years to date
	Methods of analysis	Observations of known localities for important/ sensitive plant populations
Results		Free flowering of montane species previously only recorded in vegetative state. It appears that the flush of flowering was greatest in the first year or two following reductions. Reduced winter grazing allows over-wintered buds to persist and respond to suitable conditions in spring. Increased variety of height and structure of habitats

		including flushes. Increased structure and flowering performance of cliff ledge vegetation was observed, although the more accessible parts of the habitat continue to be grazed. Flowering and seed production is likely to be a key part of adaptation of arctic-alpine species to climate change pressures. A variation in response spatially was observed, with longer swards developing at low level and more subtle structural changes at altitude. A significant reduction in sheep grazing to an annual average of around 0.5 ewes per ha with no winter grazing has allowed a number of montane in which flowering was previously suppressed to flower. This is through a combination of reduced spring grazing and lack of winter grazing allowing over-wintered buds to survive. Grazing pressure on palatable cliff ledge communities has reduced, although still selectively grazed in more accessible areas. Vegetation structure has become more variable at a range of scales, with generally taller vegetation on more productive lower slopes.
Notes	Limitations identified by author	
	Limitations identified by review team	Observational/ casual study rather than quantitative. Observations should be verified from well designed monitoring.
	Evidence gaps and/pr recommendations for further research	Repeat of original baseline including vegetation surveys and fixed point photography and targeted population monitoring of key species.
	Sources of funding	NE Internal

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	Welch, D. & Rawes, M.
	Year	1964
	Aim of study	To study the effects of removing sheep from upland grazings
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	High level upland grasslands. Not described in detail
	Eligible population	The North Pennines study area is described in terms of grazing history and human influence, along with climate and geology. Vegetation types only briefly mentioned.
	Inclusion and exclusion criteria	The exclosures were placed within areas which appeared relatively homogeneous. Presumably chosen to be representative of main grassland communities, but no system of selection presented.

	Setting	Three areas of Moor House NNR, in the N Pennines, England. Plots located at 686m on Hard Hill, 747m on Knock Fell and 823-840m on Little Dun Fell.
Methods of allocation to intervention/control	Methods of allocation	No replication. Siting of exclosures likely to have been subjective. No comparison plots at outset (1955), added in 1962.
	Intervention description	Basically grazing exclusion vs background agricultural grazing levels.
	Control/comparison description	Comparison areas (i.e. subject to background grazing) established in 1962, but no baseline established outside of exclosures at start of experiment.
	Sample sizes	Three 40 x 40m exclosures established in different parts of the reserve. Species frequency measured from a pin frame with ten pins, placed at systematically at 100 locations. Herbage sampled at 6 1.5x1m plots in each ungrazed area and grazed comparison.
	Baseline comparisons	Baseline measures from inside exclosures against which change is measured. No equivalent from grazed area.
	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Change in vegetation composition, and standing biomass of ungrazed areas.
	Secondary outcome measures	Small-scale mapping of change in small quadrats
	Follow-up periods	Exclosures in place for 7 years at point of study.
	Methods of analysis	Limited analysis using binomial sign-test (i.e. the probability that the number of increases and decreases are greater than they would be by chance). Some detailed

		mapping of fixed quadrats (25cm x 25cm) in each plot.
Results		Significant increases in ungrazed areas seen in *A tenuis, D cespitosa, D flexuosa* and *F rubra* on Knock fell, D flexuosa, F ovina and C bigelowii on Little Dun Fell. The later species however decreased on Hard Hill, where N stricta reduced by half (59 to 33 pin hits). Areas at Knock Fell mapped as dominated by J squarrosus in 1955 were dominated by D flexuosa in 1962. Flowering herbs have been markedly reduced (e.g. *M verna, T drucei* and *V myrtillus*), especially on species-rich parts of Knock Fell. However *A millefolium* had increased sharply. The exclosures had generally reduced in diversity, decreasing in mean number of species per pin frame, and each species averaging more hits per frame. The control areas each averaged more species, and fewer hits per species, from each pin frame. Sward height has increased on deeper soils within exclosures, but changed little on thin soils where some of the flowering herbs persisted. Bryophyes and lichens generally decreased. There is little evidence of new species in 1962, other than the fern *D dilitata* and moss *Plagiothecium denticulatum*. Standing crop of fine-leaved grasses was higher in enclosed plots at Knock Fell and Hard Hill, compared to annual production in the ungrazed area, but there was little difference at Little Dun Fell. Overall, the difference at Little Dun Fell was least. The litter layer increased in all exclosures compared with ungrazed plots, with greatest increase on Great Dun Fell, the highest plot where decomposition would be slowest. The total number of species in the ungrazed area on Knock Fell fell from 93 to 67 species, but there was little change at the other two sites. Seven years of grazing exclusion on three high level plots in the Northern Pennines has shown that palatable grasses increased in frequency, with reductions in mat grass and heath rush. In the most calcareous and species-rich plot low-growing herbs reduced in frequency, particularly on deeper soils where grasses grew taller. In the most species-rich exclosure the total number of species fell by one third, and very few new species were recorded at any site. The accumulation of a litter layer may have longer term

		implications for soil nutrient status and micro-organism activity.
Notes	Limitations identified by author	No grazed baseline established at start of experiment.
	Limitations identified by review team	No replication
	Evidence gaps and/pr recommendations for further research	Further studies of trajectory of change, including effects on soil processes of increased litter layer.
	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? h) What are the effects of absence or abandonment of grazing on moorland biodiversity and other ecosystem services?
Study Citation	Welch, D. & Rawes, M. (1964) The early effects of excluding sheep from high-level grasslands in the North Pennines. Journal of Applied Ecology, 1, 281-300
Study Design Category	2
Assessed by & when	D Martin 18/12/12

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: High level upland grasslands. 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 North Pennines study area is described in terms of grazing history and human influence, along with climate and geology. Vegetation types only briefly mentioned.
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: Selection not described, other than they were placed within areas which appeared relatively homogeneous. Presumably chosen to be representative of main grassland communities, but no system of selection presented.

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: No replication. Siting of exclosures likely to have been subjective. No comparison plots at outset (1955), added in 1962.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Basically grazing exclusion vs background agricultural 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: Treatments in place for seven years at point of this study. Intended that it will continue for longer
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: It is at higher end of altitudinal range of upland grasslands in England (up to 840m)
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐++	Comments: Controls reflected prevailing agricultural grazing conditions. Intervention is grazing removal – atypical, but many areas currently undergoing reductions in grazing pressure with possible local

		abandonment

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: Composition sampled in 100 systematically placed pin frames, with 10 pins each. Whilst comparisons could be made with control in 1962, there was no baseline for the control to allow change to be assessed outside of exclosure. Annual herbage production measured in fenced areas outside of main plot to allow comparison with accumulation inside exclosures. Upper herbage and stubble/ litter layer sampled.
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 no baseline for grazed plots
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: In relation to study objeictves
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: Seven years of exclosure

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐-	Comments: No comparison 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 replication
4.3 Were the estimates of effect size given or calculable?	☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	☐+	Comments: binomial sign-test (i.e. the probability that the number of increases and decreases are greater than they would be by chance).
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 change at 1% and 5% levels given for sign test of species frequency change
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: Comparative study – no replication, subjective allocation, no control at baseline.
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 are high level so may be less typical of more extensive lower level grassland, and study is low powered – no replication, limited control.

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	D Welch
	Year	1984
	Aim of study	Studies in the grazing of heather moorland in north east Scotland 1. Site descriptions and patterns of utilisation
	Study design	1
	Quality score	= QA5.1 The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made. -
	External validity	= QA5.2 The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made.
Population and setting	Source population	32 sites used in a 5000 sq km area across Aberdeenshire, Kincardineshire, Inverness-shire and Perthshire. The Western ones lay at higher altitudes in the valleys of the Feshie and Clunie (tributaries of the Spey and Dee), surrounded by the Grampians 800-1100m high. Remaining sites were in tracts of moorland having less pronounced relief, lying between the Dee and the Don.
	Eligible population	Calluna dominated moorland.
	Inclusion and exclusion criteria	Blanket bog deliberately excluded. All sites predominately Calluna dominated but with wide range of other plant species.

	Setting	Scottish moorland
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	
	Secondary outcome measures	
	Follow-up periods	
	Methods of analysis	
Results		Pages of disconnected results. 1 or 2 statements particularly relevant – 'The factors most influencing occupance were nearness to improved grasslands or swards containing many attractive graminoids and the role of each moorland tract in the management of the farm to which it belonged' 'Calluna undoubtedly experienced substantial amounts of grazing at the present sites

		and its attractiveness increased in winter compared to graminoids.
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NR

Quality Assessment Checklist: Quantitative Study 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	Studies in the grazing of heather moorland in north east Scotland 1. Site descriptions and patterns of utilisation D Welch Journal of Applied Ecology (1984) -21 pp179-195
Study Design Category	2
Assessed by & when	Alison Hiles 4/3/2013

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.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 32 sites used in a 5000 sq km area across Aberdeenshire, Kincardineshire, Inverness-shire and Perthshire. The Western ones lay at higher altitudes in the valleys of the Feshie and Clunie (tributaries of the Spey and Dee), surrounded by the Grampians 800-1100m high. Remaining sites were in tracts of moorland having less pronounced relief, lying between the Dee and the Don. Climate and soils and vegetation given in great detail.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Blanket bog deliberately excluded. All sites predominately Calluna dominated but with wide range of other plant species.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Chosen to represent different land uses and habitat types

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments: Half the sites established in June 1969 and half in June 1970. 'Each consisted of 0.4-2ha of relatively homogenous vegetation within which 8 15x1m plots were positioned for measurements of dung deposition.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: The plots were divided into groups: 1.Range restricted (fenced into an area of no more than 50 ha) with improved grassland available 2.Range restricted, no grassland 3.Unrestricted with grassland available 4.Unrestricted with no available grassland
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: Areas and their management too variable to give justifiable results.
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: All dung was first cleared from the plots. When so much dung was removed that depletion of soil nutrients could have affected plant growth, macerated dung of the species involved was returned to the plots, the amounts given at any one time kept small to prevent the vegetation becoming more attractive than elsewhere on the site. When snow lay on the ground monitoring was postponed until a substantial thaw had occurred. Farmers often moved all stock on or off the moorland. Range was restricted at many cattle-grazed sites. At some sites, supplementary feeding took place nearby.
2.5 Is the setting applicable to the UK?	☐++ ✔ ☐+ ☐-	Comments: Scotland is part of the UK but not necessarily typical for England. The dung was separated into cattle, horses, sheep, red deer, roe deer, red grouse, and lagomorphs (grouped brown and mountain hares and rabbits) which is not typical

	☐NR ☐NA	of English uplands.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measures subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++ ✔☐+ ☐- ☐NR ☐NA	Comments: Dung volumes determined by water displacement in a measuring cylinder. Collection made at 3-weekly intervals across the whole of each plot. The dung was separated into cattle, horses, sheep, red deer, roe deer, red grouse, and lagomorphs (grouped brown and mountain hares and rabbits). Monitoring of occupancy and utilisation was continued for at least 4 years. Utilisation of the main plant species present at each site estimated in 16x1sq m quadrats placed alongside the dung plots in standard position. Assessment of shoots or leaves grazed in current year's growth made 4 times a year – long shoots in Calluna and Erica, stems in Sarothamnus scoparius, leaves in Alchemilla alpina and leaves or tillers in graminoids.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	✔☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++ ☐+ ☐- ✔☐NR ☐NA	Comments:
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ✔☐+ ☐- ☐NR	Comments: Does presence of dung accurately measure 'occupancy'?

	☐NA	
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ☐+ ✔ ☐- ☐NR ☐NA	Comments: there was such a multitude of explanatory variables considered that it became obvious that the range of the study was too great for really meaningful conclusions to be drawn.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for?	☐++ ☐+ ✔ ☐- ☐NR	Comments:

Were sub-group analyses pre-specified?	☐NA	
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ✔ ☐-	Comments: The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ✔ ☐-	Comments: The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made.

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	D Welch
	Year	1984
	Aim of study	Studies in the grazing of heather moorland in north east Scotland. 2.Response of heather
	Study design	2
	Quality score	=QA 5.1 The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made. -
	External validity	=QA 5.2 The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made.
Population and setting	Source population	As Welch 1
	Eligible population	As Welch 1
	Inclusion and exclusion criteria	
	Setting	

Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	
	Sample sizes	Assessments of Calluna trend were made in a standard pattern alongside the 8 dung plots at each site. Height measured each Sept at 10 random positions in the 16x 1sq m plots. Annual extension measured on10 shoots/quadrat.
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Assessments of Calluna trend were made in a standard pattern alongside the 8 dung plots at each site. Height measured each Sept at 10 random positions in the 16x 1sq m plots. Annual extension measured on10 shoots/quadrat.
	Secondary outcome measures	
	Follow-up periods	
	Methods of analysis	
Results		'Heather declined under heavier grazing and increased mainly at sites receiving little dung.' 'The herbivores usually select for current year's growth and the biggest losses in cover, height and biomass were all less than the biggest gains. But ruminants consume some older growth and break branches by feeding and trampling.'

		'the large depositions of cattle' *(dung)* 'often killed heather, giving niches quickly colonized by herbs and graminoids, whilst viable seeds of these plants were transmitted in the dung.'
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NR

Quality Assessment Checklist: Quantitative Study 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	Studies in the grazing of heather moorland in north east Scotland. 2.Response of heather D Welch Journal of Applied Ecology (1984)– 21 pp197-207
Study Design Category	2
Assessed by & when	Alison Hiles – 5/3/2013

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.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: As Welch 1.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: As Welch 1.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments:

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Assessments of Calluna trend were made in a standard pattern alongside the 8 dung plots at each site. Height measured each Sept at 10 random positions in the 16x 1sq m plots. Annual extension measured on10 shoots/quadrat.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: : The plots were divided into groups: 1.Range restricted (fenced into an area of no more than 50 ha) with improved grassland available 2.Range restricted, no grassland 3.Unrestricted with grassland available 4.Unrestricted with no available grassland
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
2.5 Is the setting applicable to the UK?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Comments: Scotland is part of the UK but not necessarily typical for 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?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Assessments of Calluna trend were made in a standard pattern alongside the 8 dung plots at each site. Height measured each Sept at 10 random positions in the 16x 1sq m plots. Annual extension measured on10 shoots/quadrat.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments:
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	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?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

Section 5: Summary		
5.1 Are the results of the study internally valid (i.e. unbiased)? How well did the study minimise sources of bias (i.e. adjusting for potential confounders)? Were there significant flaws in the study design	☐++ ☐+ ✔ ☐-	Comments: The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made.
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐++ ☐+ ✔ ☐-	Comments: The study is so broad-based as to be virtually impossible to use to draw specific conclusions, though there are many interesting observations made.

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	D Welch
	Year	1985
	Aim of study	To evaluate the contribution that germination and colonization on dung make towards succession on heather moorland and identify differences between main herbivores.
	Study design	2 observational quantitative survey.
	Quality score	-
	External validity	+
Population and setting	Source population	Extensive upland moorland grazing with dwarf shrub communities and grazed by a range of domestic and wild herbivores.
	Eligible population	Blanket bog deliberately excluded. All sites predominately Calluna dominated but with wide range of other plant species.
	Inclusion and exclusion criteria	Chosen to represent different land uses and habitat types. Dung was collected at 12 sites in four years of 6. Selection methods not described. “several” samples of one or two herbivore types collected per site. Germination in situ on cattle, sheep and grouse

		dung examined at three sites. Surveys of plants establishing in cattle dung at six sites. Not sure how this relates to the other three sites mentioned.
	Setting	32 sites used in a 5000 sq km area across Aberdeenshire, Kincardineshire, Inverness-shire and Perthshire. The Western ones lay at higher altitudes in the valleys of the Feshie and Clunie (tributaries of the Spey and Dee), surrounded by the Grampians 800-1100m high. Remaining sites were in tracts of moorland having less pronounced relief, lying between the Dee and the Don.
Methods of allocation to intervention/control	Methods of allocation	Correlative survey type study. Half the sites established in June 1969 and half in June 1970. 'Each consisted of 0.4-2ha of relatively homogenous vegetation within which 8 15x1m plots were positioned for measurements of dung deposition
	Intervention description	Different farming grazing regimes with and without access to improved grassland.
	Control/comparison description	The plots were divided into groups: 1.Range restricted (fenced into an area of no more than 50 ha) with improved grassland available 2.Range restricted, no grassland 3.Unrestricted with grassland available 4.Unrestricted with no available grassland
	Sample sizes	Dung was collected at 12 sites in four years of 6. Selection methods not described. "several" samples of one or two herbivore types collected per site. Germination in situ on cattle, sheep and grouse dung examined at three sites. Surveys of plants establishing in cattle dung at six sites. Not sure how this relates to the other three sites mentioned. Transect counts at 7 sites of establishment of grass species not normally present in heather moorland.
	Baseline comparisons	Survey over time. No baseline as such for dung germination studies.

	Study sufficiently powered	N/A
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Frequency of occurrence and cover of different species and groups germinating on or colonising dung.
	Secondary outcome measures	Differences between species
	Follow-up periods	Dung collected four times over 6 years.
	Methods of analysis	No analysis of effects as such. Significance given for grazer species t-test.
Results		Seedlings that arose by germination from dung gained much less cover than plants colonising the deposits vegetatively. However, several species transmitted in cattle dung attained greater cover than in the previously existing vegetation e.g. *Cerastium holosteoides*, *Lolium perenne, Poa annua, Poa pratensis, Rumex acetosella, Stellaria media* and *Veronica serpyllifolia*. Surveys showed that *Anthoxanthum odoratum, Holcus lanatus, Poa annua* and *Poa pratensis* were the grasses most frequently introduced to moorland sites and increases in the number of their establishments was associated with heavy dung deposition by cattle. The contribution of dunging to the overall impact of the herbivores on the composition of the vegetation was appreciable only with cattle but the gains in cover of graminoids and herbs were less than the decline in *Calluna vulgaris* due to plant mortality below the deposits. About a quarter of the greater impact of cattle on heather compared to sheep was ascribed to dunging.
Notes	Limitations identified by author	
	Limitations identified by review team	Not clear how sites for germination studies were selected, and how dung sample was obtained – subjectivity? No attempt to correlate to explanatory variables.

	Evidence gaps and/pr recommendations for further research	Longer term effect of dunging and seed introduction on moorland communities
	Sources of funding	NR

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	Studies in the grazing of heather moorland in north east Scotland 4. Seed dispersal and plant establishment in dung D Welch Journal of Applied Ecology (1985) -22 pp461-472
Study Design Category	2
Assessed by & when	D Martin 10/3/2013

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: Part of a larger study in which 32 sites used in a 5000 sq km area across Aberdeenshire, Kincardineshire, Inverness-shire and Perthshire. The Western ones lay at higher altitudes in the valleys of the Feshie and Clunie (tributaries of the Spey and Dee), surrounded by the Grampians 800-1100m high. Remaining sites were in tracts of moorland having less pronounced relief, lying between the Dee and the Don. Climate and soils and vegetation given in great detail in first paper in series (assessed by A Hiles 4/3/2013)
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: Blanket bog deliberately excluded. All sites predominately Calluna dominated but with wide range of other plant species.
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐-	Comments: Chosen to represent different land uses and habitat types. Dung was collected at 12 sites in four years of 6. Selection methods not described. "several" samples of one or two herbivore types collected per site. Germination in situ on cattle, sheep and grouse dung examined at three sites. Surveys of plants establishing in cattle dung at six sites. Not sure how this relates to the other three sites mentioned. Transect counts at 7 sites of establishment of grass species not normally present in heather moorland.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: Not clear. Half the sites established in June 1969 and half in June 1970. 'Each consisted of 0.4-2ha of relatively homogenous vegetation within which 8 15x1m plots were positioned for measurements of dung deposition
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐NA	No real explanatory variables identified. Basically just a survey of germination. The wider sample plots were divided into groups: 1.Range restricted (fenced into an area of no more than 50 ha) with improved grassland available 2.Range restricted, no grassland 3.Unrestricted with grassland available 4.Unrestricted with no available grassland
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐+	Comments: Largish sample size. No treatments imposed as such.
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 identification or control of confounding factors, which could include access to different types of grazing, of other dietary supplements. No stirring of larger dung amounts so proportion of viable seed germinating may vary.
2.5 Is the setting applicable to the UK?	☐+	Comments: Scotland is part of the UK but not necessarily typical for England. The dung was separated into cattle, horses, sheep, red deer, roe deer, red grouse, and lagomorphs (grouped brown and mountain hares and rabbits) which is not typical of English uplands.

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measures subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐+	Comments: Mainly just observations of germination in glasshouse and in situ. Cover measured by ruler and counts made. In the glass house large plants were removed to maintain light and allow further regeneration.
3.2 Were all outcome measurements complete?	☐++	Comments:

Were all/most of the study population that met the defined study outcome definitions likely to have been identified?		
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐+ ☐-	Comments: to a point. Longer term impacts on vegetation community and role of dung in change could be further assessed.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐+	Comments: Yes.
3.5 Were there similar follow up times in exposure and comparison groups?	☐NA	Comments: Not a control/ comparison study.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐+	Comments: Yes – over 6 years

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐NR	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	No real explanatory variables included.
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: basically just frequency counts, with t-test of difference in some species between dung of different species.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐+	Comments: No analysis of effects as such. Significance given for grazer species t-test.

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: Broad range of sites but basically just a survey and no attempt to correlate to explanatory variables. Reasonable large sample and timescale so Does give an indication of the role of dung in species spread and vegetation change, and some indication of livestock species effect, although largely observational.
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: Main grazer species and habitat and vegetation species typical of upland farming areas in UK.

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	Welch, D., Scott, D. Mitchell, R. & Elston, D. A.
	Year	2006
	Aim of study	To determine the effects of reducing deer numbers within extensive sites and to record how long it took for suppressed heather to recover.
	Study design	2
	Quality score	+
	External validity	+
Population and setting	Source population	Upland heather moorland. Change in deer numbers and sheep in the Scottish context.
	Eligible population	The two study areas (glens) are representative of upland dwarf shrub heath, mainly wet heath, particularly Scottish Highlands. Vary in altitude and altitudinal range, but both high in UK terms. The areas are not in controlled burning rotation.
	Inclusion and exclusion criteria	As a correlative study, the two glens were the sampling units and systematically sampled over the whole area (90+ plots)

	Setting	Glen Derry and Glen Lui, Mar Lodge Estate, Cairngorm Mountains, Scotland. Valley bottoms at 500m and 400m respectively
Methods of allocation to intervention/control	Methods of allocation	Correlative study. Deer numbers from counts as well as estimates from dung plots
	Intervention description	Reduction in red deer numbers over time
	Control/comparison description	Two glens compared, subject to similar reductions. Winter feeding took place in one of the glens
	Sample sizes	Veg measurements and dung counts in 90+ plots in each area. Utilisation in 25 plots in one year and 60 in another. Fifteen shoots per lot.
	Baseline comparisons	All measurements made in first year. Deer dung density, lagomrph dung index and utilization lower in Glen Derry.
	Study sufficiently powered	NR
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Heather utilisation, height and cover, and change over time.
	Secondary outcome measures	Effect of soil moisture and distance from grass.
	Follow-up periods	Measured over 10 year period
	Methods of analysis	Linear mixed model using RML to analyse heather utilisation for each glen, fixed effects for year, lagomorphs dung index and deer pellet counts. Trend over time assessed against between-plot variation. Fixed effects model for red-deer pellet counts including terms for soil wetness and distance from grassland. The difference in effect of latter between glens was tested.

Results		Pellet group counts declined after the first three years then declined, with the decline much more marked in Glen Derry to 32% of the initial count, compared to 84% in Glen Lui. Rabbits were abundant or locally frequent in most years in Glen Lui, but largely absent in Glen Derry. Mountain hares were present in moderate numbers in both glens, with more dung in plots with heather. Heather shoot utilization showed similar trends and there was a significant relationship with deer dung counts in Glen Lui and lagomorphs dung index in Derry. The effect of food provision was seen in dung densities in Glen Lui. Deer dung tends to be much less on wet soils than dry and moist, although this difference reduced over time in Derry, and difference in utilization here was always small. Changes in heather cover were initially small, but mean increase was significant over time. Change was smallest and lowest in Derry. Change in height was apparent in the first 4 years in most plots, but not dry soils in Lui where heaviest utilization occurred. Highly significant increases occurred after this in Derry, but changes in Lui were small. Near the grassland in Lui utilization was highest and reflected in little height increase, with greater increases in further zones despite lower annual growth increments here. Cover increased near the grassland in a similar pattern to growth increment. Heather recovery contrasted in the two areas, with cover gains in Lui and height in Derry whilst remaining sparse. This probably reflects the main grazer, with rabbits taking shoot tips and encouraging lateral spread from buds, and deer grazing whole shoots and side branches, with trampling adding to pressure. In lightly utilized areas in Derry, the heather grows taller in the absence of rabbits, but remains sparse due to more extensive wet soils.
Notes	Limitations identified by author	Winter dieback and heather age (which can influence recovery) not measured directly.
	Limitations identified by review team	Focussed on heather and no comment on response of other species important in wet heath and related communites.

	Evidence gaps and/pr recommendations for further research	Burning trials could be carried out to assess effects on recovery
	Sources of funding	

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

Name of Evidence Review: ____Upland_____

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

Review Question	a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services?
Study Citation	Welch, D., Scott, D. Mitchell, R. & Elston, D. A. (2006). Slow recovery of Heather (*Calluna vulgaris* L. (Hull)) in Scottish moorland after easing of heavy grazing pressure from red deer (*Cervis elphus* L) Botanical Journal of Scotland 58, 1-17
Study Design Category	2
Assessed by & when	D Martin 13/12/12

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: Upland heather moorland. Change in deer numbers and sheep in the Scottish 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 two study areas (glens) are representative of upland dwarf shrub heath, mainly wet heath, particularly Scottish Highlands. Vary in altitude and altitudinal range, but both high in UK terms.
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 a correlative study, the two glens were the sampling units and systematically sampled over the whole area (90+ plots)

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++	Comments: Correlative study. Deer numbers from counts as well as estimates from dung plots
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Direct link between herbivore density and grazing pressure
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: Limited. Rabbit grazing taken into account. Effects of winter dieback and heather age mentioned in discussion, but not assessed. Soil moisture considered – affects grazing pressure and heather growth
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes, but Scottish Highlands and deer focussed

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: Good sample size for utilization measurements and cover estimates, as well as dung density counts. Annual growth increment of heather only measured in two years, in a sub-set of plots
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 shoot growth only in two years
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	☐++	Comments:
3.4 Were outcomes relevant?	☐++	Comments:

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

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: Soil moisture included, and rabbit grazing
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: Linear mixed model using RML to analyse heather utilisation for each glen, fixed effects for year, lagomorphs dung index and deer pellet counts. Trend over time assessed against between-plot variation. Fixed effects model for red-deer pellet counts including terms for soil wetness and distance from grassland. The difference in effect of latter between glens was tested.
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	☐+	Comments: Large samples, some potential confounders adjusted for (soil wetness, distance from grass) but not all (dieback, heather age).

confounders)? Were there significant flaws in the study design		
5.2 Are the findings generalisable to the wider source population (i.e. externally valid)? Are there sufficient details given to determine if the findings of can be generalised across the population (i.e. habitat, species)?	☐+	Comments: Not typical of English situation, but similar communities and deer grazing has some analogy to sheep.

Quality Assessment Checklist: Quantitative Study Experimental v2.0

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 Citation	Welch, D. (1998) Response of bilberry *Vaccinium myrtillus* L stands in the Derbyshire Peak District to sheep grazing, and implications for moorland conservation. Biological Conservation 83, 155-164.
Study Design Category	2
Assessed by & when	D Martin 21/10/12

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: Limited description of UK range, useful review of bilberry ecology
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 is sheep grazed bilberry and heather dominated moorland, but difficult for one area to represent the geographical variation of the community. The author notes that the sites are lower altitude than typical for the community (H18 *V myrtillus – D flexuosa* heath), and is transitional to H9b. (*Cladonia* sub-comm of *Calluna – D flexuosa* 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: Two exclosure systems at Ashop valley located to pick up some of the variation in the habitat – one with mixed bilberry/ heather and one of almost pure bilberry. A third site (Park Hall Moor) was established later on a different moor to give more information about seasonal sheep grazing on pure bilberry moorland. This was not subject to the same range of treatments as Ashop sites. There is likely to be a degree of subjectivity in location of study areas at both sites.

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: Summer and winter grazing treatments had two replicates at each exclosure. Not random, but one of each treatment placed uphill and downhill. The experiment ran for 6 years to minimise bias from unusual weather conditions.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Treatments are summer and winter grazing, implemented by opening and closing plots at appropriate times, year round grazing, and no grazing. Well described. Whilst it is desirable to understand the effects of winter grazing, winter only grazing is not usual in real systems.
2.3 Was the exposure to the management intervention(s) (and/or comparison(s)) adequate? Was lack of exposure sufficient to cause important bias? Consider consistency of implementation (e.g. was there unplanned variation in timing of exposures)	☐+	Comments: The experiment ran for 6 years to minimise bias from unusual weather conditions. This is likely to be adequate to allow differences to develop. Both blocks at Ashop site set up at the same time. Two grazing treatments are seasonal (summer and winter) so implemented at different times, but ran for same number of years. The winter and summer grazing periods were slightly different in later years to account for lower summer densities and keep seasonal accumulated occupancy equal. A third site was open to year-round grazing and only surveyed and counted in one year.
2.4 Was contamination acceptably low? Did any of the comparison population receive the management intervention(s) or vice versa? Was it sufficient to cause important bias?	☐++	Comments: Opportunity for contamination would be if seasonal treatments were not changed over at the right time or a plot not properly closed. There is no report of this and it is likely the treatments were applied as designed.
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: Some potential for hare or grouse to contribute to the grazing and not distinguished from sheep grazing. Likely to be minimal. No other interventions reported.
2.6 Were the wider/eligible/sample population(s)/area(s) representative of the England/UK Resource.	☐+	Comments: The site is likely to be typical of dwarf shrub heath with prominent bilberry, but difficult for one area to represent the geographical variation of the community. The author notes that the sites are lower altitude than typical for the community (H18 *V*

		myrtillus – D flexuosa heath), and is transitional to H9b. (*Cladonia* sub-comm of *Calluna – D flexuosa* heath
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: Sheep grazing is typical management for bilberry heathland. Most sites will be year-round grazed, or summer grazed where agr-environment schemes have required off-wintering. Winter only grazing is unusual, but it is desirable to understand the impacts of winter grazing. Opening up plots at start of winter may result in a flush of grazing.

Section 3: Outcomes		
3.1 Were outcome variables/measures reliable? Were outcome variables/measurements subjective or objective. How reliable were the outcome measures (e.g. inter- or intra- reliability scores, observer bias?)? Was there any indication that measures had been validated/other QA?	☐+	Comments: Main outcome measure is species frequency from point quadrats on a systematic grid. A small number of species or species groups were recorded, so mis-identification unlikely to be significant. Proportion of heather and bilberry shoots grazed was also measured in fixed areas, in percentage bands to minimise error. It is acknowledged that there may be some background grazing of hares and grouse, although impact of these discounted
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, vegetation and cover measurements made annually. Utilisation measured at 4 intervals per year.
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐++	Comments: Yes, main outcome is vegetation condition and change in 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: Dung counts used as surrogate for occupancy/ grazing pressure.
3.5 Were there similar post-treatment time intervals in exposure and comparison groups?	☐+	Comments: Both blocks at Ashop site set up at the same time. Two grazing treatments are seasonal (summer and winter) so implemented at different times, but ran for same number of years. The winter and summer grazing periods were slightly different in later years to account for lower summer densities and keep seasonal accumulated occupancy equal.

		A third site was open to year-round grazing and only surveyed and counted in one year
3.6 Was the post-treatment time interval meaningful? Was the interval long enough to assess long-term effects?	☐++	Comments: Monitored for 6 years, which is an adequate period to obtain meaningful data.

Section 4: Analyses		
4.1 Were exposure and comparison groups similar at baseline? If not, were they adjusted [in the analyses]? Were there any differences between groups in important confounders at baseline?	☐+	Comments: Reported as similar, but likely to have been a degree of variation in, for example, heather cover at the heather/ bilberry site. The two blocks were chosen to reflect slightly different starting conditions.
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 presented. Two replicates of each treatment in each block, except no grazing, which has one. Large numbers of plots in this type of experiment make it time-consuming and expensive.
4.3 Were the estimates of effect size given or calculable?	☐++	Comments: Change in % cover of key species given, and plots of change in bilberry and heather cover at each treatment at both blocks.
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: Fairly simple analysis of mean dung deposition rates and paired t-tests of between treatment differences in dung deposition, utilisation and species cover change (for utilisation data paired t-tests were carried out on individual rows of sampling grid). Line graphs of utilisation and change in heather and bilberry cover 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 levels of paired t-tests given to $p < 0.001$.
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: Limited replication, but plots likely to be fairly representative of surrounding habitat and bilberry heath elsewhere. Study ran for 6 years to account for weather effects.

bias (i.e. adjusting for potential confounders)? Were there any significant flaws in the study design?		
5.2 Are the findings generalisable to the wider source population(s)/area(s) and nationally (i.e. externally valid)? Are there sufficient details given to determine if the findings can be generalised across the population(s)/area(s) and nationally (i.e. habitat, species)?	☐+	Comments: Yes – heathland communities similar to elsewhere, although will not represent the full geographical range of the community and associated variation in production. Sheep grazing is typical management, but winter-only treatment

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	Welsh, D.
	Year	1998
	Aim of study	To investigate seasonal patterns of utilisation of bilberry by sheep and resulting effects on cover
	Study design	Non-randomised block – two sets of block exclosures with two replicated per treatment.
	Quality score	+
	External validity	+
Population and setting	Source population	UK Bilberry heathland
	Eligible population	Sheep grazed bilberry and heather dominated moorland. The author notes that the sites are lower altitude than typical for the community (H18 *V myrtillus – D flexuosa*

		heath), and is transitional to H9b. (*Cladonia* sub-comm of *Calluna – D flexuosa* heath)
	Inclusion and exclusion criteria	Two exclosure systems located to pick up some of the variation in the habitat – one with mixed bilberry/ heather and one of almost pure bilberry.
	Setting	Ashop Valley and Park Hall Moor, Derbyshire. Sites located at 310m – 350m.
Methods of allocation to intervention/control	Methods of allocation	Summer and winter grazing treatments had two replicates at each exclosure. Not random, but one of each treatment placed uphill and downhill. The experiment ran for 6 years to minimise bias from unusual weather conditions.
	Intervention description	Treatments are summer and winter grazing, implemented by opening and closing plots at appropriate times, year round grazing, and no grazing. Well described. Whilst it is desirable to understand the effects of winter grazing, winter only grazing is not usual in real systems.
	Control/comparison description	Comparison is year round grazing on open plots.
	Sample sizes	Two sets of blocks with two replicate per treatment in each block. Dung counts and utilisation on two sub-plot areas per treatment plot. Vegetation measured at 200-300 points per plot.
	Baseline comparisons	Reported as similar, but likely to have been a degree of variation in, for example, heather cover at the heather/ bilberry site. The two blocks were chosen to reflect slightly different starting conditions. Initial composition at the two blocks is presented.
	Study sufficiently powered	No power analysis presented. Two replicates of each treatment in each block, except no grazing, which has one. Large numbers of plots in this type of experiment make it time-consuming and expensive.
Outcomes and methods of analysis (inc effect	Primary outcome	Main outcome measure is species frequency from point quadrats on a systematic grid. Proportion of heather and bilberry shoots grazed was also measured, as was vegetation

size, CIs for each outcome and significance)	measures	height.
	Secondary outcome measures	Limited replication
	Follow-up periods	Monitored for 6 years, which is an adequate period to obtain meaningful data.
	Methods of analysis	Fairly simple analysis of mean dung deposition rates and paired t-tests of between treatment differences in dung deposition, utilisation and species cover change (for utilisation data paired t-tests were carried out on individual rows of sampling grid). Line graphs of utilisation and change in heather and bilberry cover over time
Results		The winter-only grazed plots had significantly greater dung deposition at the heather-bilberry site than the pure bilberry site. Seasonal effects varied between sites, with year-round grazed plots at the bilberry site having significantly greater pellet counts than seasonal plots, but the mixed heather-bilberry plots having slightly lower depositon at the year-round grazed plots than winter-grazed. The actual counts however peaked in October for the preceding eight-week period, suggesting sheep consistently chose to graze the bilberry swards much more heavily in autumn than the rest of the year. On winter-grazed plots occupancy remained high into the October-December period as almost all of the summer growth was available to graze. Shoot utilisation reflected sheep occupancy, with summer grazed plots having lower rates of utilisation, but differences were reduced through bursts of heavy usage when plots were opened in spring. The greatest increase in bilberry utilisation was recorded in August- October, and October- January for heather. Patterns of occupancy and utilisation at the hayfield site followed similar patterns to the Ashop sites. Bilberry heights changed little at the three grazing treatments, but increased in the ungrazed plot. Bilberry cover appeared unaffected by season of grazing, but crowberry appeared to benefit from winter protection. In the ungrazed plots, both bilberry and heather grew significantly taller, and the grass component and crowberry increased significantly, whilst mat grass decreased. At the heather-bilberry site heather increased in cover and

		height at the expense of bilberry cover under all grazing treatments, despite an average grazing pressure of 1.4 sheep per hectare (based on conversion of 17 pellet groups per day). Crowberry increased in cover at the winter-protected plots, and rowan saplings have appeared in the summer-protected plots. The shoot utilisation rates on heather are said to be above that which would produce biomass utilisation levels considered to be sustainable. It is postulated that conservation grazing regimes may be too cautious (DM comment – the relationship between shoots removed and sustainable utilisation rates may be subject to a number of factors and require further clarification). Spatial variation in grazing pressure is suggested as necessary to maintain heterogeneity in dwarf shrub moorland, or clear objectives need to be set as a particular stocking rate will favour certain species over others.
Notes	Limitations identified by author	The exclosure system inadequately represents what happens on the open hill as sheep can readily remove the available shoots when the plots are opened, so minimising the differences between seasons of grazing. Whether there is a different seasonal response of bilberry on the open hill remains unproven (but Hayfield site?).
	Limitations identified by review team	Limited replication. Difficulty of translating dung into stocking rates, and whether the impact of shoot utilisation rates can be compared between year round, and seasonal (i.e. biomass production may differ between the grazing regimes, resulting in different off-take for same shoot utilisation).
	Evidence gaps and/pr recommendations for further research	Studies needed on the effects of age on bilberry palatability, and how the mix of different aged stems resulting from branching affects sheep foraging. Examination of response of bilberry and heather to higher grazing pressures than at the sites in this study, to examine the effects on bilberry of stocking rates that keep heather in check, and to examine the role of burning in combination with grazing in maintaining moorland bilberry stands. More attention needs to be given to conservation of bilberry moorland and dependant fauna.
	Sources of funding	Joseph Nickerson Heather Improvement foundation 1990-93, MAFF 94-96. National

		Trust constructed the exclosures and moved fences.

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? **d**) Over what timescales can grazing-related change in plant structure and diversity be observed or expected?

Study details	Authors	Welch, D., Scott, D. & Thompson, D.B.A.
	Year	2005
	Aim of study	To investigate the effect of increased sheep grazing, as a result of snow fencing, on *Carex bigelowii- Racomitrium lanuginosim* moss heath
	Study design	Correlative study using multiple transects 2
	Quality score	+
	External validity	+
Population and setting	Source population	Montane moss-heath vegetation on high mountain plateaus in the Scottish Highlands
	Eligible population	*Carex bigelowii – Racomitrium lanuginosum* moss heath on Glas Maol, Scotland, close to and extending away from a ski-fence designed to hold snow on a ski run
	Inclusion and exclusion criteria	Selection dictated by the positioning of a snow fence, the effects of which the study is designed to evaluate.

	Setting	Mountain plateau of Glas Maol, extending above 940m to 1020m altitude
Methods of allocation to intervention/control	Methods of allocation	Transects placed at points along fence to cover range of sheep ranging from fence ends.
	Intervention description	The study measured the impact of prevailing grazing levels on vegetation
	Control/comparison description	The transects are designed to extend across a gradient of grazing, from high levels near the fence, to levels more typical of the wider plateau, which provides a comparison.
	Sample sizes	18 transects extending from the ski fence, with six 50-point quadrats per transect. Dung from 6? Plots per transect.
	Baseline comparisons	Baseline is the first year of assessment (1990), but this is four years after fence erected.
	Study sufficiently powered	No power analysis given
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Point quadrat occurrence of species, from 50 points at 6 locations on the transect, to give cover measures.
	Secondary outcome measures	none
	Follow-up periods	Monitored over a 12/13 year period
	Methods of analysis	Change in species between time periods assessed by t-tests, relationships between trend and variables tested by simple and multiple regression analysis. Significance presented to 0.001g
Results		Changes in vegetation composition close to the sheep fence where sheep concentrate are more significant over the 12 years than distant from the fence. Away from the fence *C Bigelowii*, although remaining dominant, was the species that showed the

		greatest decline, but by much less than near the fence. Agrostis increased here and two *Cladonia* species showed significant declines in this area. Adjacent to the fence there was a highly significant increases in grasses, and a highly significant decline in *C bigelowii* and *R lanuginosum*. Cover of the latter species was already one third lower near the fence than distant form it in when monitoring began (1990), four years after erection of the fence. *R lanuginosum* loss was more closely correlated to snow-lie than sheep pellet-group density, although it did decrease as pellet-group density increased. *Agrostis* increase was highly significantly related to the pellet group density, and was higher close to the fence, and at the more accessible zones near the fence ends. Beyond the plot 13-15m from the fence, dung counts indicate only a negligible decline in sheep usage, so represents the background grazing levels. Vegetation trends in 19-20m and 39-40m plots therefore represent the wider habitat and remain favourable for nesting dotterel at current grazing rates. There is indication though of on-going slow loss of lichens and bilberry, suggesting grazing-related modification, although the greatest changes took place before 96/97. The paper refutes Rodwell's hypothesis (1992b) that grazing converts moss dominate to bilberry dominated heath.
Notes	Limitations identified by author	None reported, some auto correlation of dung counts and snow lie.
	Limitations identified by review team	Little , Since data from different years on different transects were combined, may be a year effect, possible background grazing by wild herbivores not accounted for separately. Might benefit from inclusion of soil chemistry parameters
	Evidence gaps and/pr recommendations for further research	Further studies at the site to disentangle the impacts of grazing, snow-lie and N deposition.
	Sources of funding	SNH, former Scottish Office

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	
Study Citation	Welch, D., Scott, D. & Thompson, D.B.A. (2005) Changes in the composition of *Carex bigelowii – Racomitrium lanuginosum* moss heath on Glas Maol, Scotland, in response to sheep grazing and snow fencing. Biological Conservation 122, 621-631
Study Design Category	2
Assessed by & when	D Martin 12/10/12

Section 1: Population		
1.1 **Is the source population or source area well described?** e.g. Was the country, habitat and biodiversity of the area well described.	☐++	Comments: Good general description of communities, and specific description of Glas Maol plateau-vegetation, geology, climate.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: Yes – largely representative of one of the main communities C big- R lanug
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 dictated by the positioning of a snow fence, the effects of which the study is designed to evaluate.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐+	Comments: Monitoring along 18 transects extending from the fence across the grazing gradient. Botanical composition on point quadrat at fixed point on transect. Dung counts in separate fixed plots at varying distance from fence.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++	Comments: Yes – basically sheep pressure based on dung counts
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: Survey rather than experimental approach
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: Natural system so there may be topographic and environmental factors - addressed through having a number of transects. Possible effects from dung counts – nutrient removal and surveyor trampling – accounted for by separation of botanical assessment from dung counts. May be some wild herbivore grazing – not separated? Since data from different years on different transects were combined, may be a year effect. Snow lie and sheep usage are confounded
2.5 Is the setting applicable to the UK?	☐+	Comments: Yes, although habitat much less extensive, and probably more degraded, in English situation

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measure subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	☐++	Comments: Yes – objective botanical measures via point quadrats. Simple dung count measures in fixed plots
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 not monitored in every year – three main data points – 1990, 1996/7, and 2002/3 (transects 1-15 surveyed in 96 and92, 16-18 in 97, 03).
3.3 Were all important outcomes assessed?	☐+	Comments: Detailed vegetation composition. Would

Were all important positive and negative effects assessed?		have benefitted from soil chemistry measurements?
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Main outcomes relate to vegetation composition – highly relevant to aims
3.5 Were there similar follow up times in exposure and comparison groups?	☐+	Comments: Monitoring began in same year on all transects, but final surveys weren't all done in same year (two groups, one year apart). However, each survey year includes measures along the transect grazing gradient.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++	Comments: Surveyed over 12/13 year period. Even baseline data (four years from erection of fence) suggested a difference in effect along the grazing gradient, which has been on-going.

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 included. Sample size is likely to be adequate
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐+	Comments: Explanatory variables tested were sheep usage (dung counts), snow lie and altitude. No soil or other environmental data included.
4.3 Were the analytical methods appropriate? Were important differences in follow-up time and likely confounders adjusted for? Were sub-group analyses pre-specified?	☐+	Comments: t –tests of botanical data vs distance from fence. Sub groups analysed included transects near end of fence (higher usage) vs those in centre, and near and far from fence. Logistical regression analysis used to investigate vegetation trend with the explanatory variables. States that they "bore in mind that many species involved multiple testing made the chance occurrence of the 0.5 P level more likely" Not quite sure how this was accounted for
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful?	☐++	Comments: Changes in species cover at differ sample periods and distance from fence, and different locations along fence, are presented with level of significance, to 0.001. Simple and multiple regression

Were confidence intervals and or p-values for the effect estimates given or calculable?		of species against snow lie, dung and altitude presented to p=0.001
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: Long-term study, covering a period of rapid change and subsequent stabilisation. Bias minimised by multiple transects along fence covering different grazing pressures, and measuring grazing usage through dung counts rather than subjective estimates of grazing pressure. Snow lie data for later years is extrapolated and assumed to be similar than earlier years
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: Site and vegetation is well described, but largely representative of Scottish highlands.

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? 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	Welch, D & Rawes, M
	Year	1966
	Aim of study	To compare the utilisation of blanket bog vegetation under different grazing regimes, and describe the effects on vegetation composition.
	Study design	2
	Quality score	-
	External validity	+
Population and setting	Source population	High level blanket bog on deep peat, with varying amounts of heather.
	Eligible population	Areas of blanket bog around the headwaters of the Tees, North Pennines, under different farm grazing regimes.

	Inclusion and exclusion criteria	Blanket bog vegetation on peat, with a heather component is included. All areas grazed by sheep under a known regime.
	Setting	Headwaters of the River Tees, Co Durham, North Pennines. Site around 550m above sea level
Methods of allocation to intervention/control	Methods of allocation	Study plots subjectively chosen, in each of three grazing units with different grazing pressures. Observational study.
	Intervention description	The study covers plots on three management units with different overall stocking rates and grazing regimes, including off-wintering on one site.
	Control/comparison description	No control as such – three grazing regimes on different areas of blanket bog measured.
	Sample sizes	3 plots, one per site.
	Baseline comparisons	Study only ran for 14 months. Vegetation sampling shows plots differed in the cover proportions of different species.
	Study sufficiently powered	No – small sample, no replication.
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Vegetation cover in quadrats estimated by eye, from 20 quadrats. Heather biomass from 10x 1m^2 quadrats. Sheep counts twice daily in each plot were also a main measure, but explanatory variable rather than outcome.
	Secondary outcome measures	None
	Follow-up periods	Sheep counted over 14 months. Vegetation measured at one point. The approach is a census, making assumption that current grazing levels reflect past practice, which has

		influenced vegetation composition and condition.
	Methods of analysis	Mean values and standard errors
Results		The three plots varied only slightly in terms of pH, soil moisture and peat depth. Grazing pressure was inversely related to the amount of heather in the vegetation. The bog with the least heather was grazed by an average of 1 sheep per 4 acres (0.6 sheep per ha). The bog with rank heather was grazed in summer only, at less than 1 sheep per 100 acres (40 ha). It is noted that heather constancy was very similar in the plots, but cover much lower where heavily grazed. The heaviest grazed had the highest proportion of heath rush and sheep's fescue, though only a few percent each, and also of Polytrichum spp. On the lightest grazed plot with highest heather cover lichens had greatest abundance. It is noted that in the vicinity heather has different height and structure either side of fences, probably due to different grazing pressure. There is no evidence that high grazing pressure has led to a reduction in Sphagnum. Over the course of the counts, grazing peaked in the winter-grazed plots in February and March, when most snow fell, even though there were fewer sheep on the fell overall. The presence of the limestone grassland adjacent to the heavily grazed plot is ruled out as a major factor for heavy grazing in the wider area, as elsewhere vastly different grazing rates are seen on different vegetation types close to each other. It is suggested that for agricultural purposes conversion to heath rush dominated grassland (shown by the authors to support up to 1 sheep per acre (2.5 sheep/ ha)), by grazing, whilst retaining around 20% heather cover, would provide best balance of year round grazing!! (This rec aimed at improving sheep productivity from moorland)
Notes	Limitations identified by author	None
	Limitations identified by review team	Low powered observational study. Grazing pressures observed in plot counts reflect the farming regimes in place, which vary significantly. Cause and effect not investigated.

	Evidence gaps and/pr recommendations for further research	Direction of change of vegetation – is heavily grazed plot in equilibrium or heather likely to be lost
	Sources of funding	None quoted. Researchers worked for NERC

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

Name of Evidence Review: _____ Upland_____

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

Review Question	What are the effects of grazing regimes and stocking rates on the maintenance and or restoration of moorland biodiversity and ecosystem service delivery? a) What is the effect of grazing on the delivery of moorland biodiversity and other ecosystem services, including timing, frequency and regularity of grazing as well as livestock numbers, and what are the differential effects on integrated moorland ecosystem services? 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	Welch, D. & Rawes, M. (1966) The intensity of sheep grazing on high level blanket bog in Upper Teesdale. Irish Journal of Agricultural Research. 5, 185-196.
Study Design Category	2
Assessed by & when	[INSERT REVIEWER NAME & DATE]

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 high-level blanket bog on deep peat, characterised by varying amounts of heather. Described in very broad terms.
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	☐++	Comments: The study is centred on the headwaters of the Tees, and covers blanket bog vegetation varying in heather cover and grazing pressure. Fairly representative of blanket bog in the Pennines, and elsewhere in UK uplands, to a degree.
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: Chosen subjectively but to represent the three areas of different grazing pressure and history. Vegetation chosen to be uniform. All on deep peat, but one site (A) near limestone grassland. Areas are of different size.

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐-	Comments: The areas were chosen as subject to different grazing pressures, and separated by fences for decades. Only one study area per fell. Relative grazing pressures pre-judged before start of study. There may be confounding environmental factors and historic grazing pressure and vegetation condition (heather cover) are confounded.
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐+	Comments: Previous studies concluded that heather component of blanket bog vegetation had been reduced by sheep grazing. Grazing pressure is main variable included, from Census approach. Peat moisture and pH also measured.
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐+	Comments: No contamination (e.g. sheep trespass) reported. But the study is concerned with observing actual grazing levels over a period, rather than imposing defined grazing levels.
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, and plots subject to environmental factors such as weather conditions and soil factors. However at similar altitude and aspect. There is likely though to be significant confounding of grazing pressure and vegetation composition and condition and pattern e.g. limestone grassland near to one plot. No mention of burning regime – if plots have been burned and time since burning.
2.5 Is the setting applicable to the UK?	☐++	Comments: Yes – good example of internationally important UK habitat.

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: Grazing animals counted twice daily over 14-month period. Counts before and after noon, but at different times over the period to avoid bias caused by daily grazing patterns. Vegetation cover in quadrats estimated by eye. Twenty random quadrats per area is probably at the lower end of the number required to accurately reflect vegetation cover. Cover estimates will be subject to observer error. Heather biomass measured at whole quadrat level in 10 qaudrats – likely to be reliable but possible error in where heather cut above ground.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions	☐++ ☐NA	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 aims, but it is a short term study so not able to measure change in vegetation parameters.
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++	Comments: Largely – scope of study is limited.
3.5 Were there similar follow up times in exposure and comparison groups?	☐++	Comments: All plots measured over the same period.
3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐-	Comments: Only 14 month study, so really a snapshot of grazing pressure – no assessment of change, or time sequence to allow correlations

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: Small scale observational study. The sample has no power for correlative analyses.
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	☐-	Comments: No, only grazing pressure
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: Basically just mean numbers of sheep per acre, with 95% CL. The methods do not allow scope for in-depth analysis.
4.4 Was the precision of the intervention effects given or calculable? Is association meaningful? Were confidence intervals and or p-values for the effect estimates given or calculable?	☐-	Comments: No comparative statistics applied.

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: One site per grazing pressure, no replication, may be subject to bias from soil and environmental factors
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: Habitat type and grazing regimes are typical of wider northern Pennines and other upland areas, although detail will vary from site to site. Upland blanket bog is a key UK habitat of international importance.

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

Name of Evidence Review: _____Upland_____

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

Review Question	Moorland grazing and stocking rates
Study Citation	Management considerations for conserving hill areas highlighted by range analysis of hill sheep. Bryony Williams, Sean Walls, Mike Gormally, Michael Walsh & Jerome Sheahan Tearmann: Irish Journal of agri-environmental research 8, 59-76, 2011
Study Design Category	1 2
Assessed by & when	Alison Hiles 11/2/2013

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.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:216.9ha of upland and peatland in Teagase Hill sheep farm in Co. Mayo, Ireland. Blanket bog and wet heath with fragmented patches of acid grassland Aspect south-south-easterly. Altitude 14-275m
1.2 **Is the eligible population or area representative of the source population or area?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
1.3 **Do the selected habitats/flora/fauna or area represent the eligible population or area?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	☐++ ✔ ☐+ ☐- ☐NR ☐NA	Comments: Western Ireland is considerably wetter than upland England but otherwise the habitat is similar despite the generally lower altitude

Section 2: method of allocation to intervention(or comparison)		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
2.2 Was the selection of explanatory variables based on a sound theoretical basis?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
2.3 Was the contamination acceptably low? Did any of the comparison group receive the exposure? If so, was it sufficient to cause important bias?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
2.4 How well were likely confounding factors identified and controlled? Were there likely to be other confounding factors not considered or appropriately adjusted for? Was this sufficient to cause bias?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Substitute ewes tracked only when core ewes were unavailable pre-lambing either because of low body condition or were twin-bearing. 4 ewes were radio-collar tracked in each of 9 season-based sampling periods between Feb 2004 and April 2006. 1 collar failed on one occasion so 35 ranges of a single seasonal sampling period
2.5 Is the setting applicable to the UK?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Western Ireland is considerably wetter than upland England but otherwise the habitat is similar despite the generally lower altitude. The temperature extremes are much smaller than upland England and winters generally less severe

Section 3: Outcomes		
3.1 Were outcome measures and procedures reliable? Were outcome measures subjective or objective. How reliable were the outcome measures (e.g. inter- or intra-rater reliability scores)? Was there any indication that measures had been validated?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 4 Scottish Blackface ewes plus 7 substitutes were randomly selected. GPS collars used to track ewes.
3.2 Were all outcome measurements complete? Were all/most of the study population that met the defined study outcome definitions likely to have been identified?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 35/36 results mapped
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
3.4 Were outcomes relevant? Where surrogate outcome measures were used, did they measure what they set out to measure?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:
3.5 Were there similar follow up times in exposure and comparison groups?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

3.6 Was the follow up time meaningful? Was the follow-up long enough to assess long-term effects?	☐++ ☐+ ☐- ☐NR ✔ ☐NA	Comments:

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample size adequate?	☐++ ☐+ ☐- ✔ ☐NR ☐NA	Comments:
4.2 Were multiple explanatory variables considered in the analysis? Were sufficient explanatory variables considered in the analysis?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: first 3 days post release excluded. 2 separate datasets produced because the same 4 ewes were tracked in summer, autumn and winter but were often unavailable in spring (pre-lambing) so not directly comparable. Corresponding numbers of locations and days should be applied for consistency but if same number of locations are applied then the number of days varied and vice versa.
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: The analysis is extremely detailed, using well-described systems.
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: Changes in behaviour caused by fitting the collars appeared to be resolved by using dummy collars on 10% of the flock for several weeks. The same GPS collar was fitted to the same ewe when tracked, confounding ewe and collar variables. Collars were programmed to record locations at 10 minute intervals and store on board. Locations retrieved after 5 weeks DM possible confounding effects of ewe interactions. No site replication
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: Difference in climate between western Ireland and upland England – particularly temperature ranges- may alter behaviour, especially during the winter season.

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	Bryony Williams, Sean Walls, Mike Gormally, Michael Walsh & Jerome Sheahan
	Year	2011
	Aim of study	Management considerations for conserving hill areas highlighted by range analysis of hill sheep.
	Study design	1 2-DM
	Quality score	= QA5.1Changes in behaviour caused by fitting the collars appeared to be resolved by using dummy collars on 10% of the flock for several weeks. The same GPS collar was fitted to the same ewe when tracked, confounding ewe and collar variables. Collars were programmed to record locations at 10 minute intervals and store on board. Locations retrieved after 5 weeks
	External validity	=QA5.2: Difference in climate between western Ireland and upland England – particularly temperature ranges- may alter behaviour, especially during the winter season.
Population and setting	Source population	216.9ha of upland and peatland in Teagase Hill sheep farm in Co. Mayo, Ireland. Blanket bog and wet heath with fragmented patches of acid grassland Aspect south-south-easterly. Altitude 14-275m
	Eligible population	Blanket bog and wet heath with fragmented patches of acid grassland
	Inclusion and exclusion criteria	

	Setting	
Methods of allocation to intervention/control	Methods of allocation	
	Intervention description	
	Control/comparison description	
	Sample sizes	
	Baseline comparisons	
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Range analyses of sample location data used to estimate range size for each individual by season for 2 years
	Secondary outcome measures	
	Follow-up periods	
	Methods of analysis	
Results		Ranges were generally elongated across the south/southeast facing slope. Patchiness greatest in summer followed by autumn, winter then spring. Activity closely related to daylight hours. Key findings – individual range sizes consist of < 20% of the available area and 51.6% of the available area was unvisited by all 11 tracked ewes. The mean number of livestock

		per unit area may be too simplistic a management guideline. Lowest range overlap for individuals occurred between summer and winter ranges – sheep generally used different patches inside and outside the plant growing season, suggesting that different patches are more at risk of grazing-related damage in different seasons. Therefore grazing management plans should be seasonal and take vegetation condition into account. **'This study reinforces the need for stocking densities to consider what is used, not what is available, for managing areas of conservation importance'**
Notes	Limitations identified by author	Research efforts concentrated on one study site – small sample size and single study site. Suggest this is used as a pilot study.
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	Should consider sampling a larger number of sites and individuals, over all seasons and multiple years to address the issue of grazing-related damage.
	Sources of funding	Teagasc under the Walsh Fellowship Scheme

Evidence Table

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

Study details	Authors	Worral, F, Armstrong, A. & Adamson, J. K.
	Year	2007
	Aim of study	To examine the consequences of different burn regimes and absence or presence of grazing on the hydrology and soil water quality of an upland peat.
	Study design	1 Partially randomized replicated block study
	Quality score	-
	External validity	+
Population and setting	Source population	The study considers upland blanket bog. Catchment is described but the vegetation could be described in more detail – e.g. NVC, and give typical peat depth. Little context- how typical of N Pennines, English uplands etc?
	Eligible population	Plots are likely to be similar broad vegetation types as much of catchment vegetation in general (similar dominant spp quoted from Marrs et al, but there is little detail on how representative they are, or were at outset. Selection of plot locations is historic (50+ years).
	Inclusion and exclusion criteria	Using pre-existing plots rather than chosen specifically for this study. They are unusual/ untypical in that they have long-tem ungrazed/ unburned treatments, and different

		rotation burn treatments (wider area unburned). However they have been used exactly for the reason that they have been subject to long term treatments. Only three dipwells per plot, and selection of water table sampling points not described.
	Setting	Hard Hill Burning Plots, Moor House NNR, N Pennines. Altitude of 550-600m
Methods of allocation to intervention/control	Methods of allocation	Largely historical. Partially random – grazing treatments non-random for practical reasons, i.e. Half of the block open to grazing with the burning treatments randomized within each grazed or ungrazed half. Only two of the four replicates per treatment combination used in this study.
	Intervention description	Grazing is subject to background grazing of the fell and subject to change over time. No attempt made to estimate grazing levels on plots e.g. through dung counts or grazed shoots. 10 year burning cycle should mean burning in 2004, but was due to take place 2005? Twenty-year cycle was therefore half way through, and there is a long-term ungrazed treatment
	Control/comparison description	Since aim is to examine effects of grazing and burning, then control should be ungrazed unburned plots. However the paper states that grazed, unburned is the control (i.e. the prevailing management in this grazing unit).
	Sample sizes	Three dipwells per plot (two replicates) sampled fortnightly from April-Sept (from June for 10 yr burn treatment). Number of dipwells (sampling points) seems low
	Baseline comparisons	Treatments had been in place for several decades, study was measuring effects of these long-term treatments. Baseline is effectively the start of the treatments, but no data on similarity at baseline?
	Study sufficiently powered	Comments: Power of the analysis is presented and indicates high probability of a type two error for some variable/ treatment combinations: pH, absorbance, DOC for burning treatments. Only two replicates of treatment combinations.

Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Depth to water table; pH; Conductivity; DOC; absorbance
	Secondary outcome measures	Specific absorbance and character of DOC
	Follow-up periods	Treatments have been in place for several decades, allowing effects to develop. May be a difficulty with the burn rotations as the two burning treatments are at different stages of cycle, but there will be cumulative effects of previous cycles.
	Methods of analysis	Comments: data normalised to adjust for effects of different sampling days, by adjusting against control treatment means for each day. However questions exist over whether control was appropriate. Not sure whether dipwell measures included individually or mean per plot- latter prob more correct). There have been concerns expressed about the analysis, particularly whether the experiment has been incorrectly treated as a fully randomized block design affecting error testing.
Results		The removal of grazing and rotational burning appear to decrease the relative depth to the water table for the majority of sample dates. The grazed plots appear to have shallower relative water tables than ungrazed, however the average values include burned plots, compared to the grazed, unburned control, suggesting burning has a larger impact than grazing. This is verified when mean depth to water table of burned and unburned plots are compared. There are significant differences between grazing and burning treatments, but after normalisation against control treatment means burning is shown to have the greatest effect. The average effect of grazing is to decrease the water table depth by 11% , with the shallowest water table (closest to the surface) found on grazed plots which are in a 10-year burning cycle. It is postulated that increased dwarf shrub growth in ungrazed and unburned plots may lead to increased evapo-transpiration and water table draw down.

		It was found that pH did not differ significantly between grazing treatments, however the power of the analysis to detect a grazing effect was shown to be low. No significant grazing effect was found for conductivity or DOC, although there was a significant interaction between burning and grazing regimes with generally lower levels of DOC in grazed plots. The study was however limited to only presence and absence of low levels of grazing, with limited replication. The effect of grazing on water table is significant but there is little evidence of effect on other parameters measured. Water tables were shallower (i.e. nearer the surface) with grazing, said to be due to reduced vegetation development. There was no significant interaction between grazing and burning. Whilst there were no significant grazing effects on water quality parameters, there were some significant interactions, with grazing appearing to enhance the effect of frequent burning in reducing conductivity and DOC. The results do not necessarily mean that peat development is greater or DOC export less on more intensively managed plots as there may be reduced presence of peat-forming plants and DOC loss may be greater through other pathways, such as surface run-off.
Notes	Limitations identified by author	Low power (i.e. high probability of type ii error) to detect significant difference for pH and absorbance between grazing treatments and DOC for burning.
	Limitations identified by review team	Only one season, low replication, low power of some analyses – probability of type ii error. Question over control selection.
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	

Quality Assessment Checklist: Quantitative Study Experimental v2.0

bName of Evidence Review: _____Upland_____

Name of Review Sub-topic (if any): _____burning on peat_____

Review Question	What are the effects of managed burning of upland peatlands on water quality (including colouration, release of metals and other pollutants and aquatic biodiversity) and water flow (including downstream flood risk), either directly or indirectly through changes in vegetation composition and structure?
Study Citation	The effects of burning and sheep grazing on water table depth and soil water quality in a upland peat habitat. Worral, Armstrong, Adamson. J Hydrology (2007)
Study Design Category	1
Assessed by & when	David Martin 18/09/2012

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: Catchment is described but the vegetation could be described in more detail – e.g. NVC, and give typical peat depth. Little context- how typical of N Pennines, English uplands etc?
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 likely to be similar broad vegetation types as much of catchment vegetation in general (similar dominant spp quoted from Marrs et al, but there is little detail on how representative they are, or were at outset. Selection of plot locations is historic (50+ years).
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: Using pre-existing plots rather than chosen specifically for this study. They are unusual/ untypical in that they have long-tem ungrazed/ unburned treatments, and different rotation burn treatments (wider area unburned). However they have been used exactly for the reason that they have been subject to long term treatments. Only three dipwells per plot, and selection of water table sampling points not 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: Largely historical. Partially random – grazing treatments non-random for practical reasons. Only two replicates per treatment combination.
2.2 Were management intervention(s) / treatments (and/or comparison(s)) well described and appropriate? Sufficient detail to replicate? Was comparison appropriate?	☐+	Comments: Largely – grazing is subject to background grazing of the fell and subject to change over time. No attempt made to estimate grazing levels on plots e.g. through dung counts or grazed shoots. 10 year burning cycle should mean burning in 2004, but was due to take place 2005?
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: Adequate exposure –long-term treatments in place. May be spatial variation in grazing across the plots as grazed plots just open to the moor.
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: Not explicitly, but water tables etc in adjacent plots unlikely to be independent of each other? Are plots sufficiently large and dipwells placed sufficiently far from treatment edges?
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: To a point. Representative primarily of N Pennines, but there is geographical variation due to altitudinal range, topography etc. Catchment is unburned, which is untypical of many upland heath/ bog areas burned for grouse.
2.7 Did the intervention(s) or control comparison(s) reflect the usual UK practice(s)?	☐+	Comments: 10 yr rotation, grazed is probably most representative. 20yr rotation or unburned and grazed 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: Empirical measurements of water table depth and water quality parameters. Time series of measurements, but only from one summer season. Methodologies not fully described
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 one burning treatment (10 yr) measurements began later than others, and was not originally intended to include this treatment. Measurements only from one year and stopped at end Sept
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	☐+	Comments: It would seem so, although This judgement best made by someone with more expertise in water quality. though interpretation of the importance of the statistically significant differences detected in water table (e.g. for runoff generation, water quality, vegetation and rates of carbon sequestration) is hampered by the lack of data on actual water table depths (as all data presented is normalised relative to the daily average across the two grazed, unburnt 'control' plots which is described as normal for this catchment
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: Again they would appear to be, but not sure I can comment fully on what most relevant water quality measures are.
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: Treatments have been in place for several decades, allowing effects to develop. May be a difficulty with the burn rotations as the two burning treatments are at different stages of cycle, but there will be cumulative effects of previous cycles.

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: Treatments had been in place for several decades, study was measuring effects of these long-term treatments. Baseline is effectively the start of the treatments, but no data on similarity at baseline? Water table measurements normalised against control, but this is grazed, unburned plot. Arguably ungrazed, unburned would be more appropriate control.
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: Power of the analysis is presented and indicates high probability of a type two error for some variable/ treatment combinations: pH, absorbance, DOC for burning treatments. Only two replicates of treatment combinations.
4.3 Were the estimates of effect size given or calculable?	☐+	Comments: estimates of variance given for each factor and interaction, but only relative effect size is apparent due to normalisation of data. Difficult to understand magnitude of any effects.
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: data normalised to adjust for effects of different sampling days, by adjusting against control treatment means for each day. However questions exist over whether control was appropriate. Not sure whether dipwell measures included individually or mean per plot- latter prob more correct). There have been concerns expressed about the analysis, particularly whether the experiment has been incorrectly treated as a fully randomized block design affecting error testing.
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 proportion of variance explained by each factor calculated for normalised and non-normalised data
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: Only one season and summer only, low replication (two replicate blocks used), low power of some analyses – probability of type ii error. Question over control selection. Low sample density per plot. Does not take account of different stages in burning cycle. The possible confounding factors from existing vegetation differences across blocks not addressed.

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: The point is made that studies reflect the later stages of burning cycle, and does not say anything about more recently burned vegetation. Whilst effects may be related to vegetation development, different vegetation parameters are not characterised. Appears to be some positive effects of grazing and burning on DOC, but needs to be balanced with peat accumulation. Also doesn't look at other pathways such as surface run-off.

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	F Worrall and J K Adamson
	Year	2008
	Aim of study	The effect of burning and sheep grazing on soil water composition in a blanket bog: evidence for soil structural changes?
	Study design	1
	Quality score	=QA5.1 Use of multiple chemical tracers reduces the sources of bias. Multiple sampling (x18) over a year helps to adjust for potential meteorological changes.
	External validity	=QA5.2: 'The Trout Beck catchment is an 11.4 sq km blanket peat area in the headwater of the River Tees.' Above 500m. Geology described in detail. Mean temperatures and rain/snowfall detailed. Veg. Dominated by Eriophorum, Calluna vulgaris and sphagnum spp. Grazed by sheep at 0.6-1lu/ha, summer months only. No burning since 1954
Population and setting	Source population	Trout beck catchment within Moorhouse NNR. Above 500m. Geology described in detail. Mean temperatures and rain/snowfall detailed. Veg. Dominated by Eriophorum, Calluna vulgaris and sphagnum spp. Grazed by sheep at 0.6-1lu/ha, summer months only. No burning since 1954

	Eligible population	Blanket bog
	Inclusion and exclusion criteria	
	Setting	'The Trout Beck catchment is an 11.4 sq km blanket peat area in the headwater of the River Tees.'
Methods of allocation to intervention/control	Methods of allocation	4 blocks heather moorland, each split in 6, of which 3 were enclosed to prevent grazing and 3 left unfenced. Within these blocks of 3, 3 burning regimes were randomly assigned.
	Intervention description	All blocks burnt in 1954, then 3 regimes set up: no further burning; burnt every 10 years, burnt every 20 years. The 10 year burn rotation plots were due to be burnt spring 2006, so times to examine the effect of burning and grazing at the end of the 10 year burn cycle. Plan of treatments provided.
	Control/comparison description	
	Sample sizes	Multiple sampling (x18) over a year helps to adjust for potential meteorological changes.
	Baseline comparisons	All first burnt together in 1954
	Study sufficiently powered	
Outcomes and methods of analysis (inc effect size, CIs for each outcome and significance)	Primary outcome measures	Analysis of water samples for Al, Fe, Ca, Mg, K, Na, Si, fluoride, chloride, bromide, nitrate, phosphate and sulphate, ph and conductivity.
	Secondary outcome measures	Depth to water table
	Follow-up periods	Experiment started in 1954 with treatments including 20 year and 10 year burn

		rotations, grazed and ungrazed. Sampling started in April 2005 and took place on 18 occasions until April 2006
	Methods of analysis	Cations, by inductively coupled plasma optical emission spectroscopy. Anions, by ion chromatography
Results		There are significant differences in soil water composition between burning regimes but only slight differences occurred with the presence of grazing and then only in conjunction with frequency of burning. 'The results were obtained at the end of the burning cycle, implying that these changes are long-lived and may be more severe immediately after burning'
Notes	Limitations identified by author	
	Limitations identified by review team	
	Evidence gaps and/pr recommendations for further research	
	Sources of funding	NE

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 effect of burning and sheep grazing on soil water composition in a blanket bog: evidence for soil structural changes? F Worrall and J K Adamson (2008) Hydrological Processes, 22, pp2531-2541
Study Design Category	1
Assessed by & when	Alison Hiles 15/2/2013

Section 1: Population		
1.1 **Are the source population(s) or area(s) well described?** e.g. Were habitat(s) and biodiversity of the area(s) well described.	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: Trout beck catchment within Moorhouse NNR. Above 500m. Geology described in detail. Mean temperatures and rain/snowfall detailed. Veg. Dominated by Eriophorum, Calluna vulgaris and sphagnum spp. Grazed by sheep at 0.6-1lu/ha, summer months only. No burning since 1954
1.2 **Are the eligible population(s) or area(s) (the sampling frame) representative of the source population(s) or area(s)?** e..g. is the floristic diversity representative of the habitat? Were important groups under-represented?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
1.3 **Are the sampled habitats/flora/fauna or area(s) representative of the eligible population(s) or area(s)?** Was the method of selection well described? Were there any sources of bias? Were the inclusion / exclusion criteria explicit and appropriate?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:

Section 2: method of allocation to intervention(or comparison)		
2.1 method of allocation of samples to management intervention(s) (treatments) (and/or comparison(s)). How was selection bias minimised? Was allocation randomised (++)? If not randomised was significant confounding likely/not likely?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: 4 blocks heather moorland, each split in 6, of which 3 were enclosed to prevent grazing and 3 left unfenced. Within these blocks of 3, 3 burning regimes were randomly assigned.
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 blocks burnt in 1954, then 3 regimes set up: no further burning; burnt every 10 years, burnt every 20 years. The 10 year burn rotation plots were due to be burnt spring 2006,so times to examine the effect of burning and grazing at the end of the 10year burn cycle. Plan of treatments provided.
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: 3 randomly placed piezometers in each plot provided access for soil water sampling. Minute details provided of the method. A list of the cations and anions sampled is given and detailed statistical analysis provided
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: Measurements over a year complete except not possible for December and January. Sampled on a regular basis, not on a range of hydrometeorological conditions so, although it covered a year it can't be said to cover a complete range of conditions
3.3 Were all important outcomes assessed? Were all important positive and negative effects assessed by the variables/measurements used?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments: On burning there is a significant decline in Ca, Mg, Na and phosphate. The only increase post burning is for Al on the 10 year burn sites, where there is a significantly lower ph. i.e. significant differences found between burning treatments in terms of soil water composition. Presence of burning appears to exclude deeper groundwater. The effect of grazing on water table depth was less than that due to burning and the vegetation changes are not

		so extreme
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: 'The evidence is that upon burning there is greater interaction with soils, not less....Changes observed upon more frequent burning are similar to those observed after severe droughts within this catchment.' 'Alternatively, the changes in soil water composition with frequent burning could simply be a consequence of vegetational changes'
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: All first burnt together in 1954
4.2 Was the study sufficiently powered to detect an intervention effect (if one exists)? A power of 0.8 is the conventionally accepted standard. Is a power calculation present? If not, what is the expected effect size? Is the sample	✔ ☐++ ☐+ ☐- ☐NR	Comments:

size adequate?	☐NA	
4.3 Were the estimates of effect size given or calculable?	☐++ ☐+ ☐- ☐NR ☐NA	Comments:
4.4 Were the analytical methods appropriate? Were any important differences in post-treatment time and likely confounders adjusted for? Were any sub-group analyses pre-specified?	✔ ☐++ ☐+ ☐- ☐NR ☐NA	Comments:
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: Use of multiple chemical tracers reduces the sources of bias. Multiple sampling (x18) over a year helps to adjust for potential meteorological changes. Question over control selection. Low sample density per plot. Does not take account of different stages in burning cycle. The possible confounding factors from existing vegetation differences across blocks not addressed. Grazing only presence, at v low rate, or absence.
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: 'The Trout Beck catchment is an 11.4 sq km blanket peat area in the headwater of the River Tees.' Above 500m. Geology described in detail. Mean temperatures and rain/snowfall detailed. Veg. Dominated by Eriophorum, Calluna vulgaris and sphagnum spp. Grazed by sheep at 0.6-1lu/ha, summer months only.

nationally (i.e. habitat, species)?		No burning since 1954