Whither scientific debate? A rebuttal of "Contextualising UK moorland burning studies: geographical versus potential sponsorship-bias effects on research conclusions" by Brown and Holden (bioRxiv 2019; 731117)

Mark A. Ashby[a]* & Andreas Heinemeyer[b]

Author details

[a]Lancaster Environment Centre, Lancaster University, Library Avenue, Lancaster LA1 4YQ
[b] Department of Environment and Geography, Stockholm Environment Institute, University of York, York YO10 5NG
* Corresponding author: m.ashby@lancs.ac.uk
ORCID: Mark Andrew Ashby 0000-0002-8291-7490, Andreas Heinemeyer 0000-0003-3151-2466

Abstract

1. We recently published a peer-reviewed critique of the EMBER report. In a preprint response, Brown & Holden (2019) resorted to making spurious accusations of undeclared competing interests, a series of disingenuous arguments about the robustness of the EMBER results, as well as false claims of sponsorship bias. We feel that much of what they wrote falls well outside the realm of respectable scientific debate.

2. Crucially, however, Brown & Holden (2019) did not address our previous criticisms by providing a robust reanalysis of the EMBER report data that correctly accounted for site and covariate effects within the same statistical model. In our reply, we also present additional flaws which further call into question the EMBER results.

3. Brown & Holden (2019) also produced a literature review to show that the EMBER results are not out of line with the broader evidence base. However, they included papers not directly relevant to the EMBER report we criticised. Therefore, we have carried out a more accurate review. Our results indicate that the quantity and quality of available literature make it difficult to contextualise the findings of the EMBER report.

4. Finally, Brown & Holden (2019) present an error-stricken analysis of grouse moor sponsorship bias within the prescribed burning literature. Their claim that grouse moor funded research "*should be treated with extreme caution by the policy community*" goes well beyond what their data allows them to say. Not only does such a claim egregiously impugn the reputation of many scientists in the field, but it also contradicts the long-established notion that it is the quality of the science that should drive evidence-led policy.

5. *Policy Implications*. The results of the EMBER report remain unreliable. Therefore, for the time being, it should not be considered as valid evidence by policymakers. We suggest that the data from the EMBER report is reanalysed to address the shortfalls that we identify. Only then can the EMBER data and findings be used to inform upland land management policy. Also, to provide clarity to policymakers, we recommend that an independent audit into evidence reliability is carried out across the prescribed burning evidence base.

Keywords
Ecosystem services, Evidence-based policy, Experimental design, Prescribed rotational burning, the EMBER project, Upland habitats, Upland land management

1. Introduction
We welcome the response by Brown and Holden (2019) (hereafter "B&H") to our EMBER commentary (Ashby & Heinemeyer 2019). Unfortunately, however, rather than address our scientific criticisms, B&H resorted to making spurious accusations about undeclared competing interests, a series of disingenuous arguments about the robustness of the EMBER report results, as well as false and unscientific claims of sponsorship bias. In fact, we would argue that most of what B&H put forward in their response falls well outside the bounds of constructive scientific debate (i.e. a discussion about the methods, results, analyses and wider implications of a research study). Perhaps this is why B&H did not submit their response to a peer-reviewed journal despite, to our knowledge, having been invited to do so by the Journal of Applied Ecology. Regrettably, the failure to respond via peer-review means that we are left with no option but to follow suit and publish an unreviewed preprint reply as a direct response.

We have divided our reply into five sections: (i) Spurious and unscientific accusations; (ii) Disingenuous arguments; (iii) Unanswered and additional methodological criticisms; (iv) Inaccurate review to contextualise the EMBER findings; and, (v) False claims of sponsorship bias. These sections address all the key claims made by B&H.

2. Rebuttal of Brown & Holden (2019)
2.1. Spurious and unscientific accusations
In their response, B&H make several erroneous claims and insinuations about undeclared competing interests. Such unscientific allegations are not only unbefitting of two Professors but, more importantly, they also undermine our previous and forthcoming work in the eyes of the scientific community, policymakers and funding bodies. It is only for the latter reason that we have chosen to step outside the bounds of valid scientific criticism and rebut their allegations.

Firstly, B&H rightly state that there were no funding declarations on the original submission of our EMBER commentary, but in the final version we added the following grants: the Natural Environmental Research Council's (NERC) CTCD and NCEO (NERC

F14/G6/105) centres, and the UKPopNet programme (NERC R8-H12-01) (Ashby & Heinemeyer 2019). There is a simple explanation for this change: we added these grants so that one of our institutions (A. Heinemeyer, The University of York) would cover the costs of open access.

B&H subsequently question why we chose to omit funding grants that A. Heinemeyer has received from the British Association for Shooting and Conservation (BASC), the Moorland Association and the Heather Trust. Very simply, none of this funding was used to complete the EMBER commentary. For example, M. A. Ashby collated the data and wrote the first draft in his spare time while working at Natural England. The time that A. Heinemeyer devoted to the EMBER commentary was taken from his role at the Stockholm Environment Institute. Despite not using any Peatland-ES-UK funding, A. Heinemeyer listed the EMBER commentary in a Peatland-ES-UK project advisory group meeting as a related research output because it clearly fell within the scientific remit of the project, e.g., determining the impacts of different blanket bog management techniques (cutting, burning and non-intervention) on "*biodiversity, carbon, greenhouse gases and water*" (Heinemeyer 2019).

Nevertheless, B&H *may* argue that we should have declared BASC, Moorland Association and Heather Trust funding regardless of whether it was used to produce the EMBER commentary; since, in their opinion, these organisations are "*pro-burning*"[1]. Thus, by implication, such funding streams represent a financial conflict of interest. This line of argumentation might have some validity if A. Heinemeyer had received *all* his Peatland-ES-UK funding from "*pro-burning*" organisations. However, the Peatland-ES-UK project is funded by a wider group of stakeholders than those selectively cited by B&H. For instance, the research for Phase 1 of the Peatland-ES-UK project (2011-2016) was fully funded by DEFRA, with some additional PhD support from Natural England and The Moorland Association (Table 1). The outputs produced so far using Phase 1 funding include the forthcoming report by Heinemeyer *et al.* (anticipated publication 2019) and a series of associated peer-reviewed papers (Carroll *et al.* 2015; Heinemeyer *et al.* 2018; Heinemeyer & Swindles 2018; Morton & Heinemeyer 2018; Heinemeyer, Berry & Sloan 2019; Heinemeyer *et al.* 2019; Morton & Heinemeyer 2019).

In total, Phase 2 of the Peatland-ES-UK project (2017-2021) is funded by 11 separate stakeholders, with only three of these being "*pro-burning*" (Table 1). Moreover, the total percentage contribution of "*pro-burning*" funding for Phase 2 is only **21.7%** (Table 1). Why did B&H demand that we declare the funding from BASC, the Moorland Association and the Heather Trust, but not the funding from United Utilities, Yorkshire Water Services,

[1] This label is unhelpful, and one could equally make the case that water companies and conservation organisations are anti-burning and, in the case of the RSPB, anti-grouse moor (but we certainly would not criticise their research on this basis). We also question the biased assumption that funding from "*pro-burning*" groups automatically represents a conflict of interest, but funding from water companies, conservation organisations, statutory organisations and government does not. Furthermore, we disagree that BASC, the Moorland Association and the Heather Trust are inherently "*pro-burning*". Ultimately, they want the best outcomes for moorland habitat, ecosystem services and, with the exception of the Heather Trust, sporting interests. Achieving these multiple objectives is context dependent and may or may not involve burning. The current support for burning by these organisations is because they think that, in the right context, it best achieves the desired outcomes. Nevertheless, by funding the research of A. Heinemeyer, these organisations clearly acknowledge that there may be other (and possibly better) ways of managing moorlands for multiple objectives. For this they should be commended - not maligned with pejorative labels.

Yorkshire Wildlife Trust, NERC and Natural England (note, actual funding dates are different from those given by B&H, see Table 1)? Also, why were these latter funding bodies ignored when they are clearly listed as funders on the Peatland-ES-UK website (Heinemeyer 2019)? B&H openly cite the Peatland-ES-UK website: "*According to the Peatland-ES-UK project website http://peatland-es-uk.york.ac.uk, (last accessed 29 July 2019) Phase 2 of that project is currently funded by the Moorland Association (another pro-burning, gun-sports lobby group) as well as BASC*" (Brown & Holden 2019). Therefore, B&H must have seen the full list of funders when they visited the website, which means they were deliberately selective in their quote.

Our last point of contention with B&H about undeclared competing interests is based on the following statement: "*We have concerns about the potential for undeclared funding and other conflicts of interest that may entirely undermine the conclusions of some moorland burning research studies (e.g. Marrs et al., 2019a*)". In our view, the failure to declare relevant competing interests does not *automatically* undermine the conclusions of a study[2]. On the contrary, we strongly believe that all publications should be assessed on their scientific quality (e.g. rigour of experimental design, data collection, analyses and conclusions) regardless of any perceptions of bias among sponsors. At the end of the day, it should be the demonstrable quality of the science that drives evidence-led policy.

There are several additional problems with the B&H statement quoted above. Firstly, the motivations for omitting conflicts of interest cannot be deduced and could be, in many cases, due to a misunderstanding of which relationships constitute a competing interest. The latter point is conveniently demonstrated by B&H when they suggest funding from the three "*pro-burning*" groups should have been declared but fail to mention funding from the eight other funders of the Peatland-ES-UK project.

Secondly, while we would never argue against the inclusion of competing interests within scientific publications, researchers whose views fall on one side of an issue (e.g. anti-burning) may more readily dismiss studies funded by organisations that, in their mind, hold the opposing view (e.g. pro-burning), especially if the study reports contrary results (i.e. the well-known phenomenon of confirmation bias, see Nickerson 1998; Talluri *et al.* 2018). Conversely, when researchers read studies funded by groups with similar views that also report confirmatory results, they are less likely to see methodological errors (another form of confirmation bias, see Hergovich, Schott & Burger 2010). Perhaps the latter example provides an insight into why the flaws within the EMBER report have remained largely overlooked until now (but see Davies *et al.* 2016a; Davies *et al.* 2016b). Also, we certainly would not argue that studies part-funded by Natural England, the RSPB or water companies are biased towards anti-burning results.

Thirdly and finally, we would argue that significant methodological errors[3] undermine the conclusions of a research study much more than undeclared competing interests.

[2] Regarding the example of Marrs *et al.* (2019) given by B&H, R.H. Marrs initially omitted financial and non-financial competing interests arising from his association with the Heather Trust.

[3] Those that make the results unreliable, e.g., incorrect statistical analysis, measurement error and inadequate experimental design.

Consequently, if B&H have any valid scientific criticisms[4] of Marrs *et al.* (2019), then they should submit a commentary for peer-review so that the authors can respond.

2.2. Disingenuous arguments

Throughout their response, B&H engage in a series of disingenuous arguments to deflect our criticisms of the EMBER report. Firstly, in reply to our criticism of the correlative Space-for-Time approach used, B&H suggest that we ignored two recent EMBER studies in which a Before-After-Control-Impact (BACI) design was utilised (Aspray *et al.* 2017; Brown *et al.* 2019). We are glad that B&H decided to use a more robust experimental design in these studies[5]; however, they were not the focus of our commentary. Indeed, we clearly stated in our EMBER critique that we were focusing on the "*four peer-reviewed research studies relating to parts 3-6 of the main EMBER report*" (Ashby & Heinemeyer 2019). While Aspray *et al.* (2017) and Brown *et al.* (2019) may have received EMBER project funding, they were not included within the EMBER report (Brown, Holden & Palmer 2014). Maybe this is because both studies were submitted and published well outside the NERC funding period for the "*Effects of Moorland Burning on the Ecohydrology of River basins*" (EMBER) project (2009-2013) (Aspray *et al.* 2017; Brown *et al.* 2019), which is highlighted on the following webpage:

http://gotw.nerc.ac.uk/list_full.asp?pcode=NE%2FG00224X%2F1&cookieConsent=A

The studies by Aspray *et al.* (2017) and Brown *et al.* (2019) also did not directly examine the effect of prescribed burning on peatland ecosystems and, even though both studies were experimental, only Aspray *et al*. (2017) utilised a BACI design (B&H claimed both studies were BACI experiments). In short, these studies are not relevant and were not the focus of our criticisms; therefore, we find it puzzling why B&H cited them at all.

After highlighting our neglect of Aspray *et al.* (2017) and Brown *et al.* (2019), B&H cite Kinako and Gimingham (1980) in support of these studies: "*Vegetation removal as a result of fire often exposes the peat surface enhancing erosion by wind and precipitation, something which has been reported long before the EMBER project (Kinako and Gimmingham, 1980)*". This supportive citation is inappropriate because, unlike Kinako and Gimingham (1980), Aspray *et al.* (2017) and Brown *et al.* (2019) did not measure the effects of burning on soil erosion[6]. Furthermore, the study by Kinako and Gimingham (1980) was conducted within dry heath vegetation overlying mineral soils with a shallow peat layer, whereas the studies by Aspray *et al.* (2017) and Brown *et al.* (2019) were conducted within

[4] Our view is that valid scientific criticisms would include a balanced and fair examination of a studies experimental design, analysis, results and conclusions. We would consider accusations of undeclared funding and competing interest as outside the bounds of scientific debate. Consequently, such accusations should be raised with the journal in question.

[5] Despite being experimental, both studies are short-term: the study by Aspray *et al.* (2017) is conducted over two days and the study by Brown *et al.* (2019) was conducted over two months. Furthermore, Aspray *et al.* (2017) failed to account for data structure during statistical analysis (samples were nested within plot, which were nested within river), which means the statistical results presented are unreliable.

[6] These studies examined the effect of artificial (peat) and natural sediment additions on natural and artificial (mesocosms) peatland river ecosystems.

very different deep peat blanket bog ecosystems. Besides, we question the reliability of the methods used by Kinako and Gimingham (1980). For example, they hammered fifteen iron stakes (45 cm long, 1.3 cm diameter) into the soil within each plot and then measured the distance from the top of the stake to the soil surface during each visit over an 18-month period (*ibid*). It is highly likely that the peat surface moved in relation to the iron stakes due to natural soil expansion and contraction (e.g. freeze-thaw or wet-dry cycles) (see, for example, Morton & Heinemeyer 2019). This issue, combined with the lack of an unburnt control[7], means the results of Kinako and Gimingham (1980) are questionable.

In our EMBER commentary, we also highlighted that none of the papers associated with the EMBER report controlled for site (i.e. catchment) effects during data analysis (Brown *et al.* 2013; Holden *et al.* 2014; Brown *et al.* 2015; Holden *et al.* 2015). In defence, B&H cited yet another study which was not part of the main EMBER report (Brown, Holden & Palmer 2014), but used vegetation data from the EMBER study catchments (Noble *et al.* 2018)[8]. Contrary to the EMBER report (Brown, Holden & Palmer 2014), the study by Noble *et al.* (2018) *did* control for site effects during data analysis. B&H point out that we referred to Noble *et al.* (2018) within our commentary (Ashby & Heinemeyer 2019), but when we cited Noble *et al.* (2018) we did so to raise the following question: there were three authors from the main EMBER report listed as co-authors on the Noble *et al.* (2018) paper (L. E. Brown, J. Holden & S. M. Palmer), so why were site effects (and data structure) correctly controlled for in Noble *et al.* (2018) (not part of the EMBER report), but not in Brown *et al.* (2013); Holden *et al.* (2014); Brown *et al.* (2015); Holden *et al.* (2015) (all part of the EMBER report)? B&H have yet to answer this fundamentally important question.

The final disingenuous argument put forward by B&H relates to this statement: "*A&H themselves have advocated the use of geographically separate study sites to defend their own research on moorland burning*". We do indeed advocate multi-site burning studies that encompass a broad range of environmental and management conditions when the hypothesis being tested requires you to do so (e.g. comparisons of burnt versus unburnt treatments). But nowhere in our work do we advocate the separation of treatments within geographically discrete and environmentally distinct sites (as is done within the EMBER report). This is because, due to inevitable differences between sites, the geographical separation of treatments introduces large amounts of unwanted statistical noise, which in turn can introduce bias when, as within the four EMBER report studies, the treatments are not randomised; such bias can never be fully controlled for during data analysis. Therefore, to reduce noise, remove bias and isolate the relationship of interest, treatments should be replicated and randomised within each study site (i.e. the treatment-within-block design)[9]. We described this argument

[7] An unburnt control might have mitigated (to some extent) the potential problem of natural peat/soil movement (i.e. shrinkage) resulting in an apparent soil loss related to management.

[8] Noble *et al*. (2018) did use vegetation data from the plots investigated within the EMBER report (Brown, Holden & Palmer 2014). However, the EMBER report did not include any analysis of this vegetation data (*ibid*). Thus, the study by Noble *et al*. (2018) cannot be considered as part of the EMBER report.

[9] We accept that it is often difficult to achieve the treatment within block design when carrying out ecological research. In such instances, researchers should, therefore, try to control for any known and unknown between-site differences during data analysis, have more cautious hypotheses and heavily caveat their results.

within §2.4 ("*Statistical inaccuracies*") of our EMBER commentary (Ashby & Heinemeyer 2019).

2.3. Unanswered and additional methodological criticisms

2.3.1. Soil temperature measurements

B&H cite Grau-Andrés *et al.* (2019b) in support of the findings of Brown *et al.* (2015). Grau-Andrés *et al.* (2019b) measured the influence of heather burning (low burn severity, high burn severity and unburnt treatments) on soil surface temperatures (measured at 2 cm depth) in both dry heath and blanket bog during spring, summer and autumn. Similar to Brown *et al.* (2015), Grau-Andrés *et al.* (2019b) found that mean soil surface temperatures were indeed significantly lower in unburnt compared to burnt treatments in the blanket bog site, but only during summer. Also similar to Brown *et al.* (2015), Grau-Andrés *et al.* (2019b) only recorded relatively small differences in mean soil surface temperatures between burning treatments (mean differences of between 0.3 and 1ºC; see Table 3 in Grau-Andrés et al., 2019b), with many of the differences falling within the margin of error (standard errors of the mean). Conversely, Grau-Andrés *et al.* (2019b) recorded much larger differences in mean soil surface temperatures between burning treatments within the dry heath site. Thus, in our view, the findings of Grau-Andrés *et al.* (2019b) complement those of other studies which suggest that the soil surface and moss layer within blanket bog ecosystems are relatively (compared to dry heath) robust to the effects of prescribed burning (both during and post-fire) (Grau-Andrés, Gray & Davies 2017; Taylor, Levy & Gray 2017; Grau-Andrés *et al.* 2018; Grau-Andrés *et al.* 2019a), which is likely due to the higher moisture levels found within the soil and vegetation (Davies *et al.* 2010; Kettridge *et al.* 2015; Grau-Andrés *et al.* 2018; Grau-Andrés *et al.* 2019b).

Unlike Brown et al. (2015), Grau-Andrés et al. (2019b) measured soil temperature using sensors that did not have external metal thermistors (iButton™). In our EMBER commentary, we suggested that the thermistors used by Brown *et al.* (2015) were periodically exposed to the sun and this explained the large and sporadic maximum soil surface temperatures that they recorded (e.g. outliers of >50ºC). The reanalysis provided by B&H attempts to deal with some of these outliers by removing the largest 10% of temperature disturbances. However, the 10% threshold used by B&H reduces the between treatment differences in mean soil temperature to levels that are close to or within the datalogger accuracy reported in Brown *et al*. (2015) (the accuracy reported was ± 0.3 ºC). Furthermore, the 10% cut-off threshold does not identify true or false measurements but just removes values by assuming that everything above the cut-off is likely to be false and everything below it is likely to be correct. Therefore, we suggest that B&H provide further reanalysis in which false values are removed by using between plot comparisons, e.g., comparing adjacent plots and removing abrupt temperature disturbance events (i.e. rapid heating due to direct sun exposure). This reanalysis should also consider the repeated measurements taken per plot.

2.3.2. Differences between catchments (and plots)

B&H correctly note that we used the same temperature and rainfall values for two burnt and two unburnt catchments during the analysis we present in our EMBER commentary (see Table 2 and 3 in Ashby & Heinemeyer 2019). This is an artefact of the UKCP09 Met Office climate data we used which is modelled at a 5 km resolution but two burnt and two unburnt catchments used by B&H were <5 km apart. Nevertheless, B&H are wrong to claim that catchments with the same climate values are not independent. For example, imagine if climate measurements were taken at the catchment-level and that the measurements for two catchments were the same – these catchments do not suddenly lose their statistical independence. Our point is that the catchments were considered independent by B&H, so we have done the same; the fact that two burnt and two unburnt catchments are assigned the same climatic values simply means that within the available resolution of the climatic data, they are allocated the same climate values, but this does not affect their statistical independence during our analysis. Indeed, it would be wrong to reduce the sample size because of this. Moreover, the "*modelling errors*" that B&H refer to are included in the data values; their effect is to amplify the residual error variance, which means that our temperature and rainfall analysis errs on the conservative side compared to an analysis based on direct catchment measurements. Thus, our main criticisms of the EMBER report remain in place and are as follows:

1. Burnt and unburnt treatments and plots were situated in geographically separate catchments.

2. Catchments and plots differed across a range of known (and likely unknown) environmental variables that may have influenced the results.

3. Between-catchment and plot differences were not (or could not be) controlled for during statistical analysis.

4. Consequently, the results of the EMBER report are currently unreliable.

We used UKCP09 Met Office 5 km gridded climate data to make the obvious point that geographically separate catchments are always different across a range of variables (be they known or unknown). In this respect, the EMBER catchments are no different. For example, even though two pairs of catchments have the same climate based on UKCP09 Met Office 5 km gridded climate values, there are still differences in raw temperature and rainfall values between the full set of catchments (see Table 2 and 3 in Ashby & Heinemeyer 2019). Importantly, we also presented other catchment-level environmental data, such as elevation, vegetation type, underlying geology and catchment size[10]. These data indicate that: catchments were at different elevations, with three unburnt catchments being at higher elevations than any of the burnt catchments; catchments were of different sizes, with two unburnt catchments being larger than any of the burnt catchments; and, unburnt catchments supported blanket bog vegetation communities, whereas, only three out of five burnt catchments supported blanket bog vegetation communities (two of them supported dry heath

[10] These data were taken from tabulated information presented in Brown *et al*. (2013), Brown, Holden & Palmer (2014), Holden *et al*. (2015) and Noble *et al*. (2018).

vegetation) (Ashby & Heinemeyer 2019). Furthermore, the dataset used by Noble *et al.* (2018)[11] suggests that catchments (and thereby plots) differed across a range of additional environmental variables, such as atmospheric pollution, vegetation height, peat depth and bare peat (Noble *et al.* 2017a). These data are presented in Table 2 and, combined with the data presented in our EMBER commentary (Ashby & Heinemeyer 2019), emphasises our point that there are obvious environmental differences between EMBER catchments.

2.3.3. Controlling for covariates and between-catchment differences

To control for between-site differences (in this case, between-catchment differences), it is standard procedure within ecology to include site and/or any known covariates as variables during data analysis (e.g. site as a random effect and covariates as fixed effects). Crucially, the influence of site and covariates must be tested alongside treatment (e.g. burnt versus unburnt) within the *same* statistical model. This process essentially partitions the variation in the dataset accounted for by non-treatment effects, which means that the effect of the treatment can be examined in isolation. Failure to account for site and covariate effects in the *same* model can lead to unreliable results. The reason for labouring this point is because B&H clearly misunderstand that site effects and covariates must be controlled for together during data analysis, and to do so requires that they are included within the *same* statistical model used to test treatment effects. Indeed, B&H provide no supporting reanalysis of the EMBER report data in which the influence of catchment and known covariates (e.g. elevation, catchment size) are tested alongside burning impacts within the same model[12]. B&H also suggest that they adequately controlled for site effects and covariates within Brown *et al.* (2013), Brown *et al.* (2015), Holden *et al.* (2014) and Holden *et al.* (2015). We see no evidence of this. Therefore, we repeat these criticisms in detail within Boxes 1 to 4. In summary, B&H have provided no evidence to rebut our assertion that the results of the EMBER report are currently unreliable.

2.3.4. Additional methodological criticisms

Boxes 1 to 4 also contain several fresh criticisms of the EMBER report. Since these additional criticisms are described in detail within the Boxes, we only provide a summary here. Firstly, Brown *et al.* (2013), Holden *et al*. (2014), Brown *et al.* (2015) and Holden *et al*. (2015) fail to correctly account for the structure of their data during analysis (Boxes 1 to 4). Consequently, each study has artificially inflated sample sizes (i.e. pseudoreplication), which means many of the significance values reported are unreliable (i.e. they are likely to be much too low). For example, Holden *et al*. (2015) artificially increased the treatment-level sample

---

[11] Incidentally, Noble *et al*. (2018) seem to have used a similar approach to ourselves with regard to other environmental variables. For example, the NH_3, N, Nox, O_3, SO2 and Acid deposition values used as explanatory variables are measured at a 5 km resolution. However, many of the vegetation plots used as replicates by Noble *et al*. (2018) were situated <5 km apart (especially plots within the same site). Thus, Noble *et al.* (2018) repeatedly used the same NH3, N, Nox, O3, SO2 and Acid deposition values for different replicate plots (alarmingly, this applies to most of the data in both datasets). This can be seen by examining the datasets used by Noble *et al*. (2018) (see Noble *et al*. 2017a). Importantly, this does not mean that the NMDS and GLMM results presented by Noble *et al*. (2018) are unreliable.

[12] For example, in a linear mixed-effect model, site (random effect) and any covariates (fixed effects) would be entered alongside treatment (fixed effect) like so: Response variable~Treatment+Covariate1+Covariate2+(1|Site).

size of streamflow measurements by 2400% (they increased *n* from 5 to 125 per treatment). In this instance, pseudoreplication could have been avoided by averaging the individual streamflow measurements taken within each catchment or using a statistical model in which individual streamflow measurements were nested within catchment and treatment (burnt/unburnt) was specified at the catchment level.

Secondly, in three of the EMBER report studies, treatment was confounded with catchment (Holden *et al.* 2014; Brown *et al.* 2015; Holden *et al.* 2015 - catchment and treatment confounding only apples to the overland flow measurements in this latter paper). For example, Holden *et al.* (2014) made comparisons between plots of different burn ages within the Bull Clough catchment (plots burnt 2 years, 4 years and >15 years prior to the study), unburnt plots (U) within the Moss Burn catchment and plots affected by a recent (<1-year-old) wildfire (W) in the Oakner Clough catchment. Thus, W and U treatments were confounded with catchment (i.e. each treatment was in a single separate catchment). Also, Moss Burn (U plots), Bull Clough (Burnt plots) and Oakner Clough (W plots) differed across a range of catchment level environmental variables (showing clear gradients), such as mean monthly temperature (5.5, 6.3 and 7.7 ^{o}C, respectively), mean monthly rainfall (147.3, 123.6 and 117.1 mm, respectively), mean catchment elevation (664, 498 and 345.5 metres, respectively) and NVC community (M19b, H9 and M20b, respectively) (See Table 2 in Ashby and Heinemeyer, 2019). The fact that there is no catchment-level replication of U and W treatments means that between-catchment differences cannot be controlled for during data analysis. As such, B&H cannot be sure whether their comparisons between U and W, and U, W and all Burnt plots, reflect between-treatment or between-catchment differences. However, it is important to note that comparisons between plots of different burn ages within the Bull Clough catchment remain valid once the issues of pseudoreplication are resolved (Box 3).

2.4. Inaccurate review to contextualise the EMBER findings

As previously mentioned, we made it abundantly clear within our commentary that our criticisms related to the peer-reviewed publications making up parts 3-6 (i.e. the data chapters) of the main EMBER report (e.g. Brown *et al.* 2013; Brown, Holden & Palmer 2014; Holden *et al.* 2014; Brown *et al.* 2015; Holden *et al.* 2015). Therefore, it is somewhat surprising that, in their attempt to defend the EMBER findings, B&H produced a literature review which included studies on *Sphagnum*, peat exposure and peat erosion – the EMBER papers we criticised did not investigate these ecosystem properties (*ibid*). To rectify this mistake, we used the burning studies collated by B&H (see supplementary Table 4 in Brown & Holden 2019) to produce a more accurate review that contextualised the findings of the EMBER papers we actually criticised.

2.4.1. Contextualising the EMBER report findings: a systematic review

Using the PICO framework (James, Randall & Haddaway 2016), the population, intervention and comparator of the EMBER report can be defined as: Population = upland blanket bog habitats within the British Isles; Intervention = prescribed rotational burning; Comparator = an unburnt (not recently burnt) control and/or plots of different burn ages (Table 3). The outcomes investigated by the EMBER report were aquatic invertebrate biodiversity, stream water chemistry, soil temperature, soil physical and chemical properties, near-surface infiltration, macropore flow, overland flow, and water table depth (Table 3). The first stage of our review involved rejecting studies that failed to match the EMBER PICO framework (Table 3). Given that our review aimed to contextualise the findings of the EMBER report, the four peer-reviewed publications associated with the report were also excluded (Brown *et al.* 2013; Holden *et al.* 2014; Brown *et al.* 2015; Holden *et al.* 2015). Consequently, we accepted 28 studies out of the initial list of 61 studies used by B&H (Supplementary Table 1). We then read the method sections of the 28 studies in order to determine study outcomes, assess the reliability of the results[13], and establish whether burning impacts had been measured using the correct spatiotemporal context[14] (Supplementary Table 1).

The results of our review show that eleven (39%) of the 28 EMBER-related studies have potentially unreliable results (Table 4). There are also a low number of relevant studies for each of the ecosystem properties investigated by the EMBER report (Table 4). Furthermore, none of the 28 studies assessed burning impacts using the correct spatiotemporal context (Table 4). So, while the EMBER findings are in broad agreement with the relevant literature, our results indicate that the quantity and quality of relevant literature are low (Table 4). Therefore, any attempt to contextualise the findings of the EMBER report would be unreliable. The reality is that, for all ecosystem properties, the prescribed burning evidence base is weak, especially on deep peat (Glaves *et al.* 2013; Harper *et al.* 2018), which may explain the ongoing debate about burning impacts amongst the wider research community (Brown, Holden & Palmer 2016; Davies *et al.* 2016a; Davies *et al.* 2016b; Douglas *et al.* 2016; Ashby & Heinemeyer 2019; Brown & Holden 2019; Evans *et al.* 2019; Heinemeyer *et al.* 2019). It is our view that, to properly understand burning impacts on blanket bog ecosystems, we desperately need a series of long-term, catchment-scale and multi-site BACI studies (e.g. Heinemeyer *et al.* anticipated publication 2019), which, importantly, also consider alternative catchment-scale management options (Harper *et al*., 2018) and key aspects of fire ecology (e.g. fire severity, frequency and size) (Davies *et al.* 2016a).

Our review also highlights that there are several potentially unreliable studies within the prescribed burning evidence base (Supplementary Table 1). Given that we only assessed 28 studies during our EMBER-related review, the number of unreliable burning studies in the wider literature is likely to be much higher than 11. Therefore, to provide clarity to policymakers, we also recommend that an independent audit into evidence reliability is carried out across the prescribed burning evidence base. Such an audit should specifically focus on critically appraising study design and data analysis (e.g. Klimisch, Andreae & Tillmann 1997; Gormley, Pollard & Rocks 2011).

2.4.2. Additional criticisms of B&H's systematic review

We have several additional criticisms of B&H's systematic review. Firstly, by not including grey literature, B&H failed to mitigate against possible publication bias (Haddaway &

[13] Unreliable studies are those which appear to have made serious errors during data analysis that have led to model misspecification and/or pseudoreplication. Consequently, the results are currently unreliable.

[14] Measuring burning impacts before and throughout one or several burn cycles at the catchment scale (with annual, and ideally monthly, measurements taken).

Bayliss 2015; James, Randall & Haddaway 2016). Secondly, B&H provide minimal detail about their search methodology, which means that it cannot be replicated and validated (Haddaway & Verhoeven 2015). For example, was a standardised search term used? If so, how was it developed (James, Randall & Haddaway 2016)? Thirdly, B&H only used two databases (Web of Knowledge and Google Scholar) to carry out their literature search, which is perhaps why they missed several relevant studies in their original search (e.g. Armstrong *et al.* 2012; Lee *et al.* 2013b; Whitehead & Baines 2018; Morton & Heinemeyer 2019).

2.5. False claims of sponsorship bias

The analysis of grouse industry sponsorship by B&H is error-stricken and borders on pseudoscience. The most obvious flaw is that, by focussing on a narrow range of ecosystem properties (equating to 61 studies), they have ignored a significant proportion of the prescribed burning evidence base. To get a rough approximation of the burning evidence B&H ignored during their analysis, we conducted a brief literature search using Web of Science (WOS) (v.5.33) and the reference lists of three relevant literature reviews (Glaves *et al.* 2013; Heinemeyer & Vallack 2015; Harper *et al.* 2018). Firstly, we entered the following standardised search term into WOS using the advanced search function:

TS=((burn OR "fire") AND (peat* OR heath* OR moor* OR "bog" OR "mire" OR upland*) AND ("habitat management" OR "biodiversity" OR "grouse" OR bird* OR plant* OR "vegetation" OR sphagnum* OR invertebrate* OR insect* OR amphibian* OR reptile* OR mammal* OR "water quality" OR "water colour" OR "flow" OR "saturated" OR "dissolved organic carbon" OR "DOC" OR hydrolog* OR infiltrat* OR "soil" OR carbon budget* OR "carbon cycling" OR carbon flux* OR "carbon sequestration" OR carbon stock* OR "carbon storage" OR ecosystem* OR environment*))*

We developed this search term for a literature review we are currently carrying out into the effects of burning on upland ecosystem services within the British Isles. Before searching, we applied the following settings: language = English; document types = article; and, timespan = 1945-2019. After the WOS search was run (on the 27th of September 2019), we then restricted the results to all those studies conducted in England, Wales, Scotland, Northern Ireland and Ireland. These studies were subsequently reviewed by reading the title, abstract and methods to ensure that they passed the following inclusion criteria:

1. The study must be published in a peer-reviewed journal: government agency/NGO reports or M.Sc./PhD theses were omitted.

2. The study must have been conducted within the British Isles: we only considered studies conducted within England, Wales, Scotland, Northern Ireland and Ireland.

3. The study must have been predominantly conducted within the uplands: categorised as all land lying at or above 250 metres. If the study was conducted within the lab, then source material/substrate must have been collected from an upland area.

4. The impacts of prescribed burning on soil, water, carbon or biodiversity must have been explicitly investigated during the study, i.e., prescribed burning was a predictor variable and soil, water, carbon and biodiversity metrics were the response variable(s). Wildfire studies were only included if they also investigated prescribed burning impacts.

5. The study must present results from a primary empirical investigation: systematic/literature reviews were omitted.

6. The study must not have been used by B&H during their analysis of sponsorship bias.

If an article met all the inclusion criteria, it was accepted and entered into an excel database. We then examined the reference lists of three relevant literature reviews (Glaves *et al.* 2013; Heinemeyer & Vallack 2015; Harper *et al.* 2018). If a reference met all the inclusion criteria above, it was entered into the excel database. Overall, during our limited search, we found an additional 45 burning-related studies (Table A1 in Appendix A). Thus, B&H's analysis ignored at least 42% of the relevant literature. We say at least because, due to time constraints, our search ignored grey literature and only used a single literature database. Therefore, it is highly likely that the number of studies ignored by B&H is much higher.

Another significant failing of B&H's analysis is the lack of statistical independence. For example, Lee *et al.* (2013a), Marrs *et al.* (2019) and Milligan *et al.* (2018) were included as three independent data points during the analysis of bias within grouse moor industry studies. However, all three studies use the same vegetation data from the same experiment, which means they are not independent. The lack of statistical independence invalidates the Fisher's Exact Test results presented by B&H and falsifies their claims of grouse industry sponsorship bias.

A further source of error within B&H's analysis is that they classified study effects based on statements made within the paper (i.e. the conclusions) rather than on the results presented. For example, Lee *et al.* (2013a), Marrs *et al.* (2019) and Milligan *et al.* (2018) are classified by B&H as finding a positive effect of burning on *Sphagnum* growth and abundance, but the results sections of these studies indicate that burning had a mixed effect on *Sphagnum*, with the direction of the effect depending on time since burn (*ibid*). It seems odd that researchers would quantify the effects of a study based on its conclusions. Indeed, this would not be the approach taken when conducting a meta-analysis. Furthermore, the positive, negative or neutral classification of study effects is deceptive. For example, there are multiple studies in which burning has a negative effect, but the mean differences between burning treatments, while significant, are minimal (e.g. a few percentage points of vegetation cover or <1^{o}C temperature difference) (see, for example, Noble *et al.* 2017b; Noble *et al.* 2018; Grau-Andrés *et al.* 2019a; Grau-Andrés *et al.* 2019b)[15]. Such nuance is not accounted for by B&H's classification, and we wonder if such small differences are ecologically relevant.

[15] As an illustrative example, the study by Noble *et al*. (2018) recorded small but significant differences in *Sphagnum* cover (<2.5%) between burnt and unburnt plots within the condition monitoring dataset.

Finally, even if B&H's analysis were correct, it is not self-evident that that grouse industry-funded research "*should be treated with extreme caution by the policy community*". It could be that grouse industry-funded research is accurate. For instance, of the 11 studies which we classified as potentially unreliable (based on actual scientific criticism not arbitrary concepts/emotions), only one was grouse industry-funded (Clay, Worrall & Aebischer 2012), whereas, six were government agency-funded (Worrall, Armstrong & Adamson 2007; Worrall & Adamson 2008; Clay *et al.* 2009; Clay, Worrall & Fraser 2009; Clay, Worrall & Fraser 2010; Ramchunder, Brown & Holden 2013). Crucially, however, the errors committed by B&H mean that their analysis of sponsorship bias is invalid, which also means that their call for mistrust of grouse industry research goes well beyond what their data allows them to say. This oversight is particularly egregious given that they have chosen a non-peer-reviewed preprint in which to make these potentially defamatory accusations and impugn the reputation of the many scientists who receive grouse-industry funding to carry out their research. We respectfully ask that any future criticisms are both considerate and polite – objective criticism should focus on the science. We all have a difficult task in obtaining funding, but the overall aim and outcome should be robust science on which to base sound policy advice, and such science can and should be scrutinised.

3. Conclusions

In responding to B&H, we have shown that the results of the EMBER report remain unreliable. Therefore, we strongly suggest that the EMBER report, at least in its current form, is not used as a major reference for underpinning moorland burning policy in the UK. To address this issue, the authors need to publish a robust reanalysis of the EMBER report data. This analysis should fully address the issues of confounded effects and pseudoreplication. The authors may also wish to make the data publicly available to allow independent researchers to assess the reliability of the EMBER results. Only then can we determine if the conclusions are robust enough to inform upland land management policy. Lastly, to ensure that only robust and reliable science is used to inform prescribed burning policy, we advise that an independent review into evidence reliability is carried out across the entire evidence base.

Acknowledgements

We want to thank Dr Nicholas Aebischer from the Game & Wildlife Conservation Trust for his helpful comments and suggestions that have improved this article.

Author contributions

M.A. and A.H. conceived the paper; M.A. collected and analysed the data and wrote the first draft of the manuscript. Both M.A. and A.H. interpreted the results, revised the manuscript and gave final approval for submission to bioRxiv.

Funding information

No funding grants were used to complete this piece of work.

Competing interests

M. A. Ashby has provided independent ecological advice and evidence synthesis services to the Moorland Association since mid-April 2019. A. Heinemeyer has no competing interests to declare (his current funding is outlined in this publication and includes a wide range of organisations and funding agencies).

Data accessibility

All data sources used are cited within this article or amended as supplementary data (e.g. Supplementary Table 1).

References

Armstrong, A., Holden, J., Luxton, K. & Quinton, J. (2012) Multi-scale relationship between peatland vegetation type and dissolved organic carbon concentration. *Ecological Engineering,* **47,** 182-188.

Ashby, M.A. & Heinemeyer, A. (2019) Prescribed burning impacts on ecosystem services in the British uplands: A methodological critique of the EMBER project. *Journal of Applied Ecology,* **00,** 1-9.

Aspray, K.L., Holden, J., Ledger, M.E., Mainstone, C.P. & Brown, L.E. (2017) Organic sediment pulses impact rivers across multiple levels of ecological organization. *Ecohydrology,* **10,** e1855.

Brown, L.E., Aspray, K.L., Ledger, M.E., Mainstone, C., Palmer, S.M., Wilkes, M. & Holden, J. (2019) Sediment deposition from eroding peatlands alters headwater invertebrate biodiversity. *Global Change Biology,* **25,** 602-619.

Brown, L.E. & Holden, J. (2019) Contextualising UK moorland burning studies: geographical versus potential sponsorship-bias effects on research conclusions. *bioRxiv***,** 731117.

Brown, L.E., Holden, J. & Palmer, S.M. (2014) *Effects of moorland burning on the ecohydrology of river basins. Key Findings from the EMBER project.* University of Leeds, Leeds.

Brown, L.E., Holden, J. & Palmer, S.M. (2016) Moorland vegetation burning debates should avoid contextomy and anachronism: a comment on Davies *et al*.(2016). *Philosophical Transactions of the Royal Society B: Biological Sciences,* **371,** 20160432.

Brown, L.E., Johnston, K., Palmer, S.M., Aspray, K.L. & Holden, J. (2013) River ecosystem response to prescribed vegetation burning on blanket peatland. *Plos One,* **8,** e81023.

Brown, L.E., Palmer, S.M., Johnston, K. & Holden, J. (2015) Vegetation management with fire modifies peatland soil thermal regime. *Journal of Environmental Management,* **154,** 166-176.

Carroll, M.J., Heinemeyer, A., Pearce-Higgins, J.W., Dennis, P., West, C., Holden, J., Wallage, Z.E. & Thomas, C.D. (2015) Hydrologically driven ecosystem processes determine the distribution and persistence of ecosystem-specialist predators under climate change. *Nature Communications,* **6,** 7851.

Clay, G.D., Worrall, F. & Aebischer, N.J. (2012) Does prescribed burning on peat soils influence DOC concentrations in soil and runoff waters? Results from a 10 year chronosequence. *Journal of Hydrology,* **448,** 139-148.

Clay, G.D., Worrall, F., Clark, E. & Fraser, E.D. (2009) Hydrological responses to managed burning and grazing in an upland blanket bog. *Journal of Hydrology,* **376,** 486-495.

Clay, G.D., Worrall, F. & Fraser, E.D. (2009) Effects of managed burning upon dissolved organic carbon (DOC) in soil water and runoff water following a managed burn of a UK blanket bog. *Journal of Hydrology,* **367,** 41-51.

Clay, G.D., Worrall, F. & Fraser, E.D. (2010) Compositional changes in soil water and runoff water following managed burning on a UK upland blanket bog. *Journal of Hydrology,* **380,** 135-145.

Davies, G.M., Kettridge, N., Stoof, C.R., Gray, A., Ascoli, D., Fernandes, P.M., Marrs, R., Allen, K.A., Doerr, S.H. & Clay, G.D. (2016a) The role of fire in UK peatland and moorland management: the need for informed, unbiased debate. *Phil. Trans. R. Soc. B,* **371,** 20150342.

Davies, G.M., Kettridge, N., Stoof, C.R., Gray, A., Marrs, R., Ascoli, D., Fernandes, P.M., Allen, K.A., Doerr, S.H. & Clay, G.D. (2016b) The peatland vegetation burning debate: keep scientific critique in perspective. A response to Brown *et al*. and Douglas *et al*. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **371,** 20160434.

Davies, G.M., Smith, A.A., MacDonald, A.J., Bakker, J.D. & Legg, C.J. (2010) Fire intensity, fire severity and ecosystem response in heathlands: factors affecting the regeneration of *Calluna vulgaris*. *Journal of Applied Ecology,* **47,** 356-365.

Douglas, D.J., Buchanan, G.M., Thompson, P. & Wilson, J.D. (2016) The role of fire in UK upland management: the need for informed challenge to conventional wisdoms: a comment on

Davies *et al*.(2016). *Philosophical Transactions of the Royal Society B: Biological Sciences,* **371,** 20160433.

Evans, C.D., Baird, A.J., Green, S.M., Page, S.E., Peacock, M., Reed, M.S., Rose, N.L., Stoneman, R., Thom, T.J., Young, D.M. & Garnett, M.H. (2019) Comment on: "Peatland carbon stocks and burn history: Blanket bog peat core evidence highlights charcoal impacts on peat physical properties and long-term carbon storage," by A. Heinemeyer, Q. Asena, W. L. Burn and A. L. Jones (Geo: Geography and Environment 2018; e00063). *Geo: Geography and Environment,* **6,** e00075.

Glaves, D., Morecroft, M., Fitzgibbon, C., Owen, M., Phillips, S.K. & Leppitt, P. (2013) *The effects of managed burning on upland peatland biodiversity, carbon and water*. Natural England evidence review NEER004, Peterborough, UK.

Gormley, A., Pollard, S. & Rocks, S. (2011) Guidelines for Environmental Risk Assessment and Management: Green Leaves III. DEFRA & Cranfield University, UK.

Grau-Andrés, R., Davies, G.M., Gray, A., Scott, E.M. & Waldron, S. (2018) Fire severity is more sensitive to low fuel moisture content on *Calluna* heathlands than on peat bogs. *Science of the Total Environment,* **616,** 1261-1269.

Grau-Andrés, R., Davies, G.M., Waldron, S., Scott, E.M. & Gray, A. (2019a) Increased fire severity alters initial vegetation regeneration across *Calluna*-dominated ecosystems. *Journal of Environmental Management,* **231,** 1004-1011.

Grau-Andrés, R., Gray, A. & Davies, G.M. (2017) *Sphagnum* abundance and photosynthetic capacity show rapid short-term recovery following managed burning. *Plant Ecology & Diversity,* **10,** 353-359.

Grau-Andrés, R., Gray, A., Davies, G.M., Scott, E.M. & Waldron, S. (2019b) Burning increases post-fire carbon emissions in a heathland and a raised bog, but experimental manipulation of fire severity has no effect. *Journal of Environmental Management,* **233,** 321-328.

Haddaway, N.R. & Bayliss, H.R. (2015) Shades of grey: two forms of grey literature important for reviews in conservation. *Biological Conservation,* **191,** 827-829.

Haddaway, N.R. & Verhoeven, J.T.A. (2015) Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. *Ecology and Evolution,* **5,** 4451-4454.

Harper, A.R., Doerr, S.H., Santin, C., Froyd, C.A. & Sinnadurai, P. (2018) Prescribed fire and its impacts on ecosystem services in the UK. *Science of the Total Environment,* **624,** 691-703.

Heinemeyer, A. (2019) Welcome to Peatland-ES-UK: Blanket Bog (upland peat) Management and Ecosystem Services in the UK. University of York, York, UK. http://peatland-es-uk.york.ac.uk/

Heinemeyer, A., Asena, Q., Burn, W.L. & Jones, A.L. (2018) Peatland carbon stocks and burn history: Blanket bog peat core evidence highlights charcoal impacts on peat physical properties and long-term carbon storage. *Geo: Geography and Environment,* **5,** e00063.

Heinemeyer, A., Berry, R. & Sloan, T.J. (2019) Assessing soil compaction and micro-topography impacts of alternative heather cutting as compared to burning as part of grouse moor management on blanket bog. *Peerj,* **7,** e7298.

Heinemeyer, A., Burn, W.L., Asena, Q., Jones, A.L. & Ashby, M.A. (2019) Response to: Comment on "Peatland carbon stocks and burn history: Blanket bog peat core evidence highlights charcoal impacts on peat physical properties and long-term carbon storage" by Evans *et al*. (Geo: Geography and Environment 2019; e00075). *Geo: Geography and Environment,* **6,** e00078.

Heinemeyer, A. & Swindles, G.T. (2018) Unraveling past impacts of climate change and land management on historic peatland development using proxy-based reconstruction, monitoring data and process modeling. *Global Change Biology,* **24,** 4131-4142.

Heinemeyer, A. & Vallack, H.W. (2015) *Literature review on: potential techniques to address heather dominance and help support 'active' Sphagnum supporting peatland vegetation on blanket* *peatlands and identify practical management options for experimental testing*. University of York draft report to Defra and Natural England, York, UK.

Heinemeyer, A., Vallack, H.W., Morton, P.A., Pateman, R., Dytham, C., Ineson, P., McClean, C., Bristow, C. & Pearce-Higgins, J.W. (anticipated publication 2019) *Restoration of heather-dominated blanket bog vegetation on grouse moors for biodiversity, carbon storage, greenhouse gas emissions and water regulation: comparing burning to alternative mowing and uncut management (Appendix by Richard Lindsay). Final Report to Defra on Project BD5104* Stockholm Environment Institute at the University of York, York, UK.

Hergovich, A., Schott, R. & Burger, C. (2010) Biased Evaluation of Abstracts Depending on Topic and Conclusion: Further Evidence of a Confirmation Bias Within Scientific Psychology. *Current Psychology,* **29,** 188-209.

Holden, J., Palmer, S.M., Johnston, K., Wearing, C., Irvine, B. & Brown, L.E. (2015) Impact of prescribed burning on blanket peat hydrology. *Water Resources Research,* **51,** 6472-6484.

Holden, J., Wearing, C., Palmer, S., Jackson, B., Johnston, K. & Brown, L.E. (2014) Fire decreases near-surface hydraulic conductivity and macropore flow in blanket peat. *Hydrological Processes,* **28,** 2868-2876.

James, K.L., Randall, N.P. & Haddaway, N.R. (2016) A methodology for systematic mapping in environmental sciences. *Environmental Evidence,* **5,** 7.

Kettridge, N., Turetsky, M.R., Sherwood, J.H., Thompson, D.K., Miller, C.A., Benscoter, B.W., Flannigan, M.D., Wotton, B.M. & Waddington, J.M. (2015) Moderate drop in water table increases peatland vulnerability to post-fire regime shift. *Scientific Reports,* **5**.

Kinako, P. & Gimingham, C. (1980) Heather burning and soil erosion on upland heaths in Scotland. *Journal of Environmental Management,* **10,** 277-284.

Klimisch, H.J., Andreae, M. & Tillmann, U. (1997) A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. *Regulatory Toxicology and Pharmacology,* **25,** 1-5.

Lee, H., Alday, J.G., Rose, R.J., O'Reilly, J. & Marrs, R.H. (2013a) Long-term effects of rotational prescribed burning and low-intensity sheep grazing on blanket-bog plant communities. *Journal of Applied Ecology,* **50,** 625-635.

Lee, H., Alday, J.G., Rosenburgh, A., Harris, M., McAllister, H. & Marrs, R.H. (2013b) Change in propagule banks during prescribed burning: a tale of two contrasting moorlands. *Biological Conservation,* **165,** 187-197.

Marrs, R.H., Marsland, E.L., Lingard, R., Appleby, P.G., Piliposyan, G.T., Rose, R.J., O'Reilly, J., Milligan, G., Allen, K.A., Alday, J.G., Santana, V., Lee, H., Halsall, K. & Chiverrell, R.C. (2019) Experimental evidence for sustained carbon sequestration in fire-managed, peat moorlands. *Nature Geoscience,* **12,** 108-112.

Milligan, G., Rose, R.J., O'Reilly, J. & Marrs, R.H. (2018) Effects of rotational prescribed burning and sheep grazing on moorland plant communities: Results from a 60-year intervention experiment. *Land Degradation & Development,* **29,** 1397-1412.

Morton, P.A. & Heinemeyer, A. (2018) Vegetation matters: Correcting chamber carbon flux measurements using plant volumes. *Science of the Total Environment,* **639,** 769-772.

Morton, P.A. & Heinemeyer, A. (2019) Bog breathing: the extent of peat shrinkage and expansion on blanket bogs in relation to water table, heather management and dominant vegetation and its implications for carbon stock assessments. *Wetlands Ecology and Management,* **27,** 467-482.

Nickerson, R.S. (1998) Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology,* **2,** 175-220.

Noble, A., Palmer, S.M., Glaves, D.J., Crowle, A., Brown, L. & Holden, J. (2017a) Data from: Prescribed burning, atmospheric pollution and grazing effects on peatland vegetation composition. University of Leeds, Research Data Leeds Repository. http://archive.researchdata.leeds.ac.uk/234/

Noble, A., Palmer, S.M., Glaves, D.J., Crowle, A., Brown, L.E. & Holden, J. (2018) Prescribed burning, atmospheric pollution and grazing effects on peatland vegetation composition. *Journal of Applied Ecology,* **55,** 559-569.

Noble, A., Palmer, S.M., Glaves, D.J., Crowle, A. & Holden, J. (2017b) Impacts of peat bulk density, ash deposition and rainwater chemistry on establishment of peatland mosses. *Plant and Soil,* **419,** 41-52.

Ramchunder, S.J., Brown, L.E. & Holden, J. (2013) Rotational vegetation burning effects on peatland stream ecosystems. *Journal of Applied Ecology,* **50,** 636-648.

Talluri, B.C., Urai, A.E., Tsetsos, K., Usher, M. & Donner, T.H. (2018) Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence. *Current Biology,* **28,** 3128-3135.e3128.

Taylor, E.S., Levy, P.E. & Gray, A. (2017) The recovery of *Sphagnum capillifolium* following exposure to temperatures of simulated moorland fires: a glasshouse experiment. *Plant Ecology & Diversity,* **10,** 77-88.

Whitehead, S.C. & Baines, D. (2018) Moorland vegetation responses following prescribed burning on blanket peat. *International Journal of Wildland Fire,* **27,** 658-664.

Worrall, F. & Adamson, J. (2008) The effect of burning and sheep grazing on soil water composition in a blanket bog: evidence for soil structural changes? *Hydrological Processes: An International Journal,* **22,** 2531-2541.

Worrall, F., Armstrong, A. & Adamson, J. (2007) The effects of burning and sheep-grazing on water table depth and soil water quality in a upland peat. *Journal of Hydrology,* **339,** 1-14.

Tables

Table 1. All the funders that have contributed to Phase 1 and 2 of the Peatland-ES-UK research frameworks. Phase 1 took place between 2012-2016, whereas Phase 2 started in 2017 and finishes in 2021.

	Phase 1		Phase 2	
Organisation	Sign date	Contribution (%)	Sign date	Contribution (%)
a) Main project funding				
Department for Environment, Food and Rural Affairs	08/2011	99.5		
Moorland Association			12/2017	3.2
Moorland Association			09/2018	12.7
Yorkshire Water Services			10/2018	33.9
United Utilities[1]			09/2018	11.9
Yorkshire Wildlife Trust			10/2018	1.3
British Association for Shooting and Conservation			10/2018	4.0
Ecological Continuity Trust (BES)			12/2018	0.5
Heather Trust			09/2019	1.8
b) PhD support funding				
Moorland Association	04/2015	0.3		
NERC iCASE			10/2017	14.2
Natural England (CASE partner)	02/2015	0.2	10/2017	0.6
NERC iCASE			10/2018	14.2
Yorkshire Wildlife Trust (CASE partner)			10/2018	0.6
c) Student support funding				
YCCSA (summer school intern for two months)			07/2016	0.3
YCCSA (summer school intern for two months)			07/2016	0.3
University of York (intern for three months)			12/2016	0.5

[1] United Utilities' contribution is set to increase in 2019 with outstanding funding TBC in April 2020

Table 2. Additional environmental data for each of the EMBER catchments. This data is taken from Noble *et al*. (2017a), and Appendix A describes how the data was calculated. Combined with the data we presented in Ashby and Heinemeyer (2019), the data in this table verifies the assertion that there are clear environmental differences between the EMBER catchments.

Management/ catchment	NH_3 deposition	N deposition	Nox deposition	O_3 deposition	SO_2 deposition	Acid deposition	Peat depth (cm)	Veg height (cm)	Bare peat (%)
Burnt catchments:									
Bull Clough	2.88	33.60	9.38	1040.36	1.91	2.51	150.00	13.75	1.89
Rising Clough	3.63	29.54	13.90	959.00	2.26	2.21	115.67	25.83	0.18
Woo Gill	0.65	21.84	5.90	1052.65	1.44	1.87	110.67	17.50	2.58
Great Eggleshope beck	1.24	17.36	5.22	679.78	1.15	1.38	150.00	27.92	0.77
Lodgegill Sike	0.58	20.44	4.82	698.12	1.26	1.68	86.92	27.08	0.03
Mean ± SEM	1.80 ± 0.62	24.56 ± 3.02	7.84 ± 1.72	885.98 ± 82.08	1.60 ± 0.21	1.93 ± 0.20	122.65 ± 12.18	22.42 ± 2.85	1.09 ± 0.50
Unburnt catchments:									
Crowden Little Beck	0.78	24.64	5.69	1034.73	1.37	2.06	128.75	12.92	1.28
Green Burn	1.90	25.48	5.81	356.89	1.20	2.02	129.25	22.92	0.00
Moss Burn	0.55	18.90	4.30	775.93	0.90	1.55	141.50	22.92	2.13
Oakner Clough	0.82	23.52	5.33	845.96	1.24	1.95	137.83	17.92	4.31
Trout Beck	1.13	29.40	5.59	393.37	1.09	2.41	132.08	20.00	0.00
Mean ± SEM	1.04 ± 0.23	24.39 ± 1.69	5.34 ± 0.27	681.38 ± 132.12	1.16 ± 0.08	2.00 ± 0.14	133.88 ± 2.50	19.34 ± 1.86	1.54 ± 0.80

Table 3. The scope of the EMBER report (Brown *et al.* 2013; Brown, Holden & Palmer 2014; Holden *et al.* 2014; Brown *et al.* 2015; Holden *et al.* 2015) expressed using the PICO framework (James, Randall & Haddaway 2016).

Population
Upland blanket bog habitats within the British Isles.

Intervention
Prescribed rotational burning.

Comparator
An unburnt (not recently burnt) control and/or plots of different burn ages.

Outcomes:
Aquatic invertebrate biodiversity, stream water chemistry, soil temperature, soil physical and chemical properties, near-surface infiltration, macropore flow, overland flow, and water table depth.

Table 4. Summary statistics from the systematic literature review of the five ecosystem properties relevant to EMBER report papers (Brown *et al.* 2013; Brown, Holden & Palmer 2014; Holden *et al.* 2014; Brown *et al.* 2015; Holden *et al.* 2015).

Ecosystem property	**Number of relevant papers**	**Number of studies presenting potentially unreliable results**	**Number of studies using the incorrect spatiotemporal context**	**Number of studies reporting a response (studies with potentially unreliable results)**	**Number of studies reporting no response (studies with potentially unreliable results)**
Prescribed burning alteration to mean and/or maximum soil temperature	3	0	3	3 (0)	0 (0)
Prescribed burning alteration to aquatic invertebrate diversity	2	2	2	2 (2)	0 (0)
Prescribed burning alteration to peat physical and/or chemical properties	15	8	15	12 (8)	3 (0)
Prescribed burning alteration to stream chemistry	7	1	7	6 (1)	1 (0)
Prescribed burning alteration to peatland hydrological functions	5	3	5	4 (3)	1 (0)
Total (note, this is not a simple sum as some papers measured more than one ecosystem property)	28	11	28	23 (11)	5 (0)

Boxes

Box 1. A detailed methodological description (black text) and critique (blue text) of Brown *et al.* (2013). This study relates to part six of the main EMBER report (Brown, Holden & Palmer 2014).

<u>Brown et al. (2013): River ecosystem response to prescribed vegetation burning on blanket peatland</u>

<u>Aim:</u> To improve understanding about the effects of rotational burning on aquatic macroinvertebrate community structure and composition within peatland streams.

<u>Study design:</u>
Comparisons were made between five burnt and five unburnt independent river catchments. Treatments (burnt versus unburnt) were located within separate catchments and were not randomly assigned to a site. Furthermore, the average distance between treatment catchments was 76.7 ± 10.9 km (See Appendix S1 in Ashby & Heinemeyer 2019) and each catchment differed across a range of known environmental variables (covariates), such as elevation, underlying geology, catchment size, vegetation community, temperature, rainfall, atmospheric pollution, peat depth, vegetation height and amount of bare peat (Table 2; Brown *et al.* 2013; Brown, Holden & Palmer 2014; Noble *et al.* 2018; Ashby & Heinemeyer 2019). It is highly likely that catchments also differed across a multitude of other unknown variables, such as fire and grazing history. Moreover, unburnt catchments were, in general, at higher elevations and wetter (Ashby & Heinemeyer 2019). Given the known and likely unknown differences between catchments, the authors should have caveated their results.

<u>Sampling strategy:</u>
Invertebrate biodiversity and benthic particulate organic matter (POM) data were collected from each river during six visits over 20 months: spring (March/April), summer (June) and autumn (Sept/Oct) in 2010 and 2011. During each visit, invertebrates and POM were sampled by collecting five random Surber samples from riffle habitats in each river (five samples per river per season). We have two criticisms of this sampling strategy. Firstly, this short-term assessment is unlikely to capture the long-term impacts of burning (rather, it more likely captures short-term effects or catchment-specific events). Secondly, it does not separate burning impacts from generic between-catchment environmental differences (e.g. climate).

Water quality data were collected from each river during five visits over 17 months: summer (June) and autumn (Sept/Oct) in 2010, and spring (March/April), summer (June) and autumn (Sept/Oct) in 2011. During each visit, water chemistry was measured by collecting water samples from within each river. The authors do not make it clear how many water samples were taken from each river per visit. Also, the authors do not say where in the river the sample was taken from or whether the sample location was randomly selected.

<u>Data analysis:</u>
Multivariate analysis of variance (MANOVA) was used to test the effects of management (burnt verses unburnt), season and their interaction on invertebrate metrics and POM (which was divided into fine and coarse: FPOM and CPOM, respectively). Individual Surber samples were treated as replicates during MANOVA analysis: this is pseudoreplication, whereby sample size (n) has been artificially inflated from $n = 1$ to $n = 5$ for each catchment visit. By artificially inflating the number of samples, the authors have reduced the error and increased the chance of finding a statistically significant effect. The authors should have either pooled the five samples per visit (by calculating mean or median values) or used a nested MANOVA in which sample was nested within season. In addition, however, the seasonal visits represent repeated measures within a catchment, so the authors should have nested season within catchment whether or not they assumed individual Surber samples were replicates. In an attempt to adjust for bias generated by differences between catchments, known covariates describing catchment characteristics should have also been included in the model, mindful that with only 8 degrees of freedom (df) for error for the burnt/unburnt comparison, the scope for including covariates is limited.

The potential for bias arising from differences between catchments means that the results cannot solely be attributed to differences in burning management (burnt versus unburnt). Furthermore, the failure to correctly account for data structure (i.e. wrongly specified error df) leads to pseudoreplication which means that the MANOVA levels of significance reported within the paper are unreliable. Somewhat surprisingly, the authors state that "*Differences between individual rivers were not assessed with MANOVA as the main focus of the study was on management effects*". This statement represents a fundamental misunderstanding of experimental design and statistical analysis because in the absence of treatment randomisation it is crucial to determine what differences exist between catchments as these have the potential to bias the results and render them uninterpretable.

Box 1. Continued. A detailed methodological description (black text) and critique (blue text) of Brown *et al.* (2013). This study relates to part six of the main EMBER report (Brown, Holden & Palmer 2014).

Data analysis:

Non-Metric Multidimensional Scaling (NMDS) was used to ordinate mean macroinvertebrate abundance data (log10 x+1 transformed) and river water chemistry metrics for each river/season combination. The NMDS analysis tested for differences between burnt and unburnt catchments in terms of invertebrate biodiversity and water chemistry. We assume this means that, as for the MANOVA, the individual Surber samples were used as the basis for analysis, leading again to pseudoreplication. For the statistical comparison of burnt versus unburnt catchments, the analysis should have been done at the catchment level (n = 5 per treatment; n = 10 in total).

One-way Analysis of Similarity (ANOSIM) was used to determine if differences in invertebrate community composition were greater between than within burnt and unburnt catchments. The authors provide no information about how the data is treated within this analysis. However, given the mistakes made during the MANOVA and NMDS analyses, it is likely that the individual Surber samples were used as the basis for analysis, again giving rise to pseudoreplication.

Box 2. A detailed methodological description (black text) and critique (blue text) of Brown *et al.* (2015). This study relates to part five of the main EMBER report (Brown, Holden & Palmer 2014).

Brown *et al.* (2015): Vegetation management with fire modifies peatland soil thermal regime

Aim: To assess the role of prescribed burning on the soil thermal regime within peatland ecosystems.

Study design:
Soil temperatures were monitored within plots of different burn ages in the Bull Clough catchment (plots burnt 2 years, 4 years, 7 years and >15 years before the study; B2, B4, B7 and B15+, respectively) and unburnt plots (U) within the Oakner Clough catchment. Three 400 m^2 plots were used for each burn age treatment (B2, B4, B7 and B15+), and these were positioned in top, middle, and bottom hillslope positions. Conversely, no information is provided about the U plots used within Oakner Clough. The authors provide no information about the number and location of U plots in Oakner Clough. Did the number and stratification of U plots match that used for burnt plots within Bull Clough (one plot per burn treatment at top, middle and bottom slope positions)? If so, then the authors would have had a choice of three U plots per hillslope position to choose from. Was the single U plot used at each hillslope position selected at random to avoid bias? The authors should have clarified this. Also, the U treatment was confounded with catchment (i.e. it was in a single separate catchment), but burn age (B2, B4, B7 and B15+) treatments were not (they were all within the same catchment). This is problematic because Bull Clough (B2, B4, B7 and B15+ plots) and Oakner Clough (U plots) differed across a range of environmental variables, such as mean monthly temperature (6.33 and 7.71 ºC, respectively), mean monthly rainfall (123.56 and 117.11 mm, respectively), mean catchment elevation (498 and 345.5 metres, respectively) and NVC community (H9 and M20b, respectively) (See Table 2 in Ashby & Heinemeyer 2019). Consequently, burnt plots (B2, B4, B7 and B15+) and U plots are also likely to have differed from each other across these environmental variables (See, for example, Fig.6a in Ashby & Heinemeyer 2019). Despite the confounding of the U treatment with catchment and the environmental differences between catchment and plots, the authors do not caveat their results.

Sampling strategy:
Soil temperature was measured within each 400 m^2 plot at four depths: 0-1 cm (the peat surface), 5 cm, 20 cm and 50 cm. Gemini PB-5001 thermistors interfaced with Gemini Tinytag TGP-4520 dataloggers were used to take soil measurements at the four different depths. The accuracy of these thermistor probes is ± 0.3 ºC "*across the temperature range experienced in this study*". Surface soil temperatures were taken by placing the thermistor "*horizontally within the top 1 cm of the peat-litter complex and checked for position every three weeks*". Sub-surface measurements were taken by inserting thermistor probes "*into the previously undisturbed soil using a wooden dowel (which was then subsequently retrieved) so that the probe tip containing the thermistor element was located at the desired measurement depth*". Soil temperatures were recorded within each plot and depth every 15 minutes for 18 months (28th of March 2010 to the 24th of October 2011). However, a wildfire on the 9th of April 2011 damaged the U plots within Oakner Clough. Therefore, "*when comparing between Bull Clough and Oakner Clough records, only those data collected pre-wildfire were used in this study to avoid any confounding effects*". It appears that they only used one thermistor within every study plot at each soil depth (thus, three thermistors for each treatment/depth combination). We have several questions about the sampling strategy. Firstly, the use of a single U catchment indicates that treatment U is confounded with catchment, so the use of U as a control means that the disturbance values used in the analysis represented differences between catchments not necessarily differences between treatments. Secondly, by locating burnt plots within a single catchment, the authors vastly reduce the reliability of generalisations made from their results. Thirdly, how did the authors choose where to insert thermistors within each plot? Differences in vegetation (height and composition) and peat depth at the point where the thermistor is inserted are likely to influence soil temperatures (as well as the water tables monitored by Holden *et al.* 2015). How did the authors try to minimise the influence of such confounding factors? This is particularly important given that only one thermistor was used per plot/depth combination. Finally, the method of thermistor insertion used to measure surface soil temperatures suggests to us that there was periodic exposure of some of the thermistors to the sun, especially given that the thermistor sensor is approximately 12 cm long. This assertion is backed up by the fact that the level of difference between maximum temperatures recorded between burnt plots (B2, B4, B7 and B15+) is considerably larger at the soil surface than at lower measurement depths (See Table 1 in Brown *et al.* 2015). Indeed, even though fire effects were also notable at 5 cm depth, the magnitude of difference was far lower than at the surface, with mean temperature increases of 0.3ºC in B2 compared to B15+ (which is within the range of measurement error reported in the paper).

Box 2. Continued. A detailed methodological description (black text) and critique (blue text) of Brown *et al.* (2015). This study relates to part five of the main EMBER report (Brown, Holden & Palmer 2014).

Data analysis:

The authors summarised the raw 15-min soil temperature data by calculating daily mean, maximum and minimum statistics for each plot/depth combination. All subsequent analysis used these descriptive statistics. The authors did not use hypothesis tests to compare daily descriptive statistics between treatments, but they did report mean values for the burn age treatments within Bull Clough (but not for the unburnt treatments within Oakner Clough) (See Table 1 in Brown *et al.* 2015). To examine if vegetation removal using fire significantly modified peatland soil thermal regime, the authors used a generalised regression approach that attempted to "*predict soil temperature time-series at burned plots (Bull Clough) from those measured at the unburned site (Oakner Clough)*". The temperature dataset was split into equal length odd and even day datasets: odd day temperatures were used to build models, whereas even day datasets were used to evaluate model performance. After the model performance had been tested, odd day models for each slope position/depth were used to predict soil temperatures at equivalent depths and slope positions in B2, B4, B7 and B15+ plots. ANOVA with Tukey HSD post hoc comparisons was then used to examine differences in average model estimates of daily mean and daily maximum random disturbance measurements between each burn age category (B2, B4, B7 and B15+ plots). In this latter analysis, it seems that the only explanatory variable entered into the ANOVA model was burn age treatment (as a fixed effect). The predictive regression model used by the authors included a parameter to account for differences between catchments (the α parameter in equation 1). Nevertheless, doing so does not avoid the issue that treatment U is confounded with catchment (i.e. U plots were all located within a single catchment). Thus, the model results and the subsequent ANOVA comparisons of daily maximum random disturbance may simply reflect catchment-specific differences between Bull Clough and Oakner Clough that are unrelated to burning; any conclusions drawn cannot be considered reliable because of this. Also, the ANOVA model used by the authors to compare average model estimates of daily mean and daily maximum random disturbance measurements ignores the data structure. Although the Methods lack detail in this area, it seems likely that daily temperatures from each plot/depth combination across the study period served as replicates; if so, these would have been repeated measures, and the ANOVA model should have been nested accordingly. However, this was not done and therefore means that the authors artificially inflated the number of replicates per treatment by a huge amount (i.e. pseudoreplication). Furthermore, the authors failed to account for hillslope position within the ANOVA analysis. In fact, they openly admit to disregarding hillslope as a factor during data analysis. A correct analysis taking into account both repeated measures and hillslope position during ANOVA analysis would yield an F-statistic with 3, 6 degrees of freedom. The erroneous analysis presented in the paper, with inflated degrees of freedom, means that the significant differences reported in Table 5 of Brown *et al.* (2015) are questionable.

Box 3. A detailed methodological description (black text) and critique (blue text) of Holden *et al.* (2014). This study relates to parts three and four of the main EMBER report (Brown, Holden & Palmer 2014).

Holden *et al*. (2014): Fire decreases near-surface hydraulic conductivity and macropore flow in blanket peat

Aim: To investigate the impact of time since burn, wildfire and no burning on surface infiltration rate, saturated hydraulic conductivity and macropore flow within peatland ecosystems.

Study design:
Comparisons were made between plots of different burn ages within the Bull Clough catchment (plots burnt 2 years, 4 years and >15 years prior to the study; B2, B4 and B15+, respectively), unburnt plots (U) within the Moss Burn catchment and plots affected by a recent (<1-year-old) wildfire (W) in the Oakner Clough catchment. Three 400 m^2 plots were used for each treatment, and these were positioned in top, middle and bottom hillslope positions. Firstly, the authors do not clarify whether the three W plots and three U plots were selected at random for each hillslope position (they selected one of three U and W plots for each hillslope position, but how did they avoid bias during plot choice?). Secondly, W and U treatments were confounded with catchment (i.e. each treatment was in a single separate catchment), and all burn treatments (B2, B4 and B15+) were within a single different catchment. Thirdly, Moss Burn (U plots), Bull Clough (B2, B4 and B15+ plots) and Oakner Clough (W plots) differed across a range of catchment level environmental variables (showing clear gradients), such as mean monthly temperature (5.5, 6.3 and 7.7 ^{o}C, respectively), mean monthly rainfall (147.3, 123.6 and 117.1 mm, respectively), mean catchment elevation (664, 498 and 345.5 metres, respectively) and NVC community (M19b, H9 and M20b, respectively) (See Table 2 in Ashby & Heinemeyer 2019). Consequently, burnt plots (B2, B4 and B15+), U plots and W plots are also likely to have differed from each other across a range of environmental variables (See, for example, Fig.5a in Ashby & Heinemeyer 2019). Finally, the authors did provide caveats for the confounding of catchment and treatment: "*We recognize that three nearby sites were used for this study and that there are likely to be between-sites differences that may not be the result of fire effects. The potential between-sites effects should be borne in mind when interpreting the results*".

Sampling strategy:
Data collection took place during July and August 2011. Within each of the three 400 m^2 plots used per treatment (top, middle and bottom slope plots), infiltration measurements were taken at six or seven random sampling points using a tension disc infiltrometer. This gave 20 samples per treatment. However, some of the samples were rejected due to rapid changes in ambient weather conditions. Thus, the number of samples used per treatment during data analysis varied between 16 to 19. It appears that individual infiltrometer samples served as replicates. Data from the tension disc infiltrometer experiments were used to calculate the infiltration rate, saturated hydraulic conductivity and macropore flow for each sample. Firstly, because measurements were taken over two months during a single year, it provides only a very limited short-term view of burning impacts on peatland infiltration rate, saturated hydraulic conductivity and macropore flow. Also, additional plots (and catchments) could have been included to increase statistical power and generalisability. Furthermore, it is unclear whether infiltration experiments were carried out during standardised weather conditions across treatments (or even on the same day). This may have been difficult to achieve given that treatments were in three different catchments, which, again, raises the issue of confounding external factors influencing the results.

Data analysis:
Infiltration rate, saturated hydraulic conductivity and macropore flow data were log10 transformed to remove skewness and heterogeneous variation within groups so that parametric tests of difference could be applied. Initially, a nested ANOVA was used to investigate the effect of burning treatment (unburnt, B2, B4, B15+ and wildfire) on infiltration rate, saturated hydraulic conductivity and macropore flow. Within the nested ANOVA model, treatment was nested within hillslope position (top, middle and bottom). However, the authors removed this nested random effect because "*Slope was not a significant factor within the nested ANOVA at p<0.05*". Thus, a one-way ANOVA was used to test the effect of treatment on infiltration rate, saturated hydraulic conductivity and macropore flow. Post-hoc Tukey's multiple comparison tests were then used to examine which individual treatments differed significantly. By pooling individual infiltrometer samples across plots, by dropping slope, by ignoring catchment and by using a one-way ANOVA to analyse treatment, the authors completely ignored the structure of their data. This led to large-scale pseudoreplication, which artificially inflated error degrees of freedom. Consequently, the significance levels presented within the paper are too low, giving rise to spuriously significant effects. In actual fact, the structure is that of a unbalanced split-plot design with subsamples, whereby treatment is specified at two levels: a top level (Burn, U, W) with no replication (1 catchment only for each), and an unbalanced lower level (B2, B4, B15+) within the Burn catchment only. At the top level, there are 3 catchments so 2 df in total, but treatment requires 2 df, leaving no df for error – it is impossible to test for differences between Burn, U and W. The lower level treatment may be analysed using a nested ANOVA applied to the Burn catchment only, with individual infiltrometer samples nested within plot and hillslope position used as a blocking factor. The lack of replication for U and W at the catchment level and the unbalanced twin-level specification of treatment preclude any meaningful comparison between B2, B4, B15+ and U, W.

Box 4. A detailed methodological description (black text) and critique (blue text) of Holden *et al.* (2015). This study relates to part four of the main EMBER report (Brown, Holden & Palmer 2014).

Holden et al. (2015): Impact of prescribed burning on blanket peat hydrology

Aim: To examine the effect of burning management on water table depth (WTD), overland flow (OFLW) and streamflow (SF) within peatland ecosystems.

Study design:
Comparisons of SF measurements were made between five burnt and five unburnt independent river catchments. Within each catchment, the authors took measurements from a single second-order stream. As highlighted in Box 1 above and by Ashby and Heinemeyer (2019), it was not possible to randomise treatments (burnt/unburnt) across catchments. Consequently, the study design used by Holden *et al.* (2015) potentially introduced bias due to known (e.g. catchment-level differences in elevation, climate, vegetation, geology, catchment size) and unknown external factors. Because of the potential bias, the authors should have caveated their results.

WTD was compared between plots of different burn ages (plots burnt 2 years, 4 years, 7 years and >10 years prior to the study; B2, B4, B7 and B10+, respectively) within five burnt catchments and unburnt plots within five additional catchments (ten catchments in total: five burnt and five unburnt). Twelve 400 m^2 soil plots were selected within each of the ten catchments. In burnt catchments, "*three replicates of each age class were chosen, with one each located in top, middle, and foot slope positions*". Conversely, in unburnt catchments, "*12 patches were selected randomly with four replicates per slope position*". WTD depth was also compared between all burnt plots and all unburnt plots across every catchment. Compared to burnt plots (B2, B4, B7 and B10+), U plots tended to be at higher elevations, on steeper slopes and more north-facing (See Figures 3 & 4 in Ashby & Heinemeyer 2019). B2, B4, B7 and B10+ also varied in terms of slope and aspect (*ibid*). The authors failed to control for all these confounding factors during data analysis. They also failed to caveat their results in light of the significant limitations afforded by their study design.

OFLW was compared between plots of different burn ages in the Bull Clough catchment (plots burnt 2 years, 4 years, 7 years and >15 years before the study; B2, B4, B7 and B15+, respectively) and unburnt plots (U) within the Oakner Clough catchment. Twelve 400 m^2 soil plots were selected within each catchment. In Bull Clough, this meant there were three replicates for each burn age located at top, middle and bottom slope positions. In Oakner Clough, this meant there were four U plots at each hillslope position. The U treatment was confounded with catchment (i.e. it was in a single separate catchment), and all the burn age (B2, B4, B7 and B15+) treatments were within the same catchment. As highlighted in Box 3 above, this is problematic because Bull Clough (B2, B4, B7 and B15+ plots) and Oakner Clough (U plots) differed across a range of environmental variables (See Table 2 in Ashby & Heinemeyer 2019). Consequently, burnt plots (B2, B4, B7 and B15+) and U plots will also have differed from each other across a range of environmental variables (See, for example, Fig.6a in Ashby & Heinemeyer 2019). Despite having only two catchments, one burnt and one unburnt, and despite the environmental differences between catchments and plots, the authors did not caveat their results.

Sampling strategy:
SF metrics were derived from rainfall and river stage data. Within each catchment, rainfall was recorded using a single automated tipping bucket rain gauge attached to Gemini Tiny Tag TGPR-1201 event recorder (Gemini Data Loggers, Chichester, UK). River stage was recorded within each second-order stream using a single Druck PDCR 1830 pressure transducer attached to Campbell CR1000 data logger that scanned at five-second intervals and recorded 15 min averages. Rainfall and river stage data were collected between March 2010 and November 2011 (20-month period). We only have one criticism of the sampling strategy. How did the authors decide where to locate the rain gauges and the in-stream pressure transducers to ensure representative measurements?

WTD was measured using a single dipwell that was inserted within each 400 m^2 soil plot (installed at approximately the centre of plot). Measurements were taken manually every three weeks over an unspecified time period. The authors provide no information about how long measurements were taken for. However, Figure 2. in Holden *et al.* (2015) suggests measurements were taken between July 2010 and October 2011. Also, the vegetation within the centre of each plot is likely to have differed, and this may have impacted upon WTD within the dipwells, yet no further detailed information is given.

OFLW was measured using ten crest-stage tubes per 400 m^2 soil plot (arranged in a 2 x 5 grid, with 1 m between the two grid lines), with measurements being taken every visit. As with WTD measurements, the authors provide no information about how long the measurements were taken for (e.g. how many visits? And over what period?). Also, where were the crest-stage tube grids located within each plot and how did the authors decide where to locate them to avoid bias?

Box 4. Continued. A detailed methodological description (black text) and critique (blue text) of Holden *et al.* (2015). This study relates to part four of the main EMBER report (Brown, Holden & Palmer 2014).

Data analysis:

During SF data analysis, 25 randomly sampled storms were used per catchment, with storm response metrics (e.g. recession time, hydrograph intensity and storm runoff ratio) being calculated for each of these storm events. This gave 125 storm responses per treatment (burnt versus unburnt; 250 storm responses in total). Differences in storm response metrics were then compared between burnt and unburnt catchments using Student's *t*-tests. By using 25 storms per catchment (n = 125 per treatment and n = 250 overall), the authors artificially inflated the number of replicates (*i.e.* pseudoreplication). The true number of replicates was n = 5 per treatment and n = 10 in total. Pseudoreplication could have been avoided by pooling the 25 storm responses per catchment (e.g. by averaging them) or using a statistical model in which storm response was nested within catchment and treatment specified at the catchment level. Moreover, the authors could have tried to control for known external variables during data analysis using covariates, which might have helped account for some of the variation between catchments. The failure to recognize the true sample size for the burnt versus unburnt comparison means that standard errors are five times smaller than they should be, leading to spuriously significant differences.

Firstly, WTD data were pooled and then compared between all unburnt and all burnt plots (across all ten catchments) using a Student's *t*-test. Additionally, differences in WTD data between plots of different burn ages (B2, B4, B7, B10+ and U plots), slope positions (top, middle and bottom) and sample dates were examined using a mixed-effects general linear model. In this model, sample date was a random factor and burn age and slope position were fixed factors. Pairwise differences between burn age and slope position treatments were calculated using Tukey's method. The authors provide no information about what served as a replicate during the overall comparisons (burnt versus unburnt plots) and comparisons between plots of different burn ages (comparisons between B2, B4, B7, B10+ and U plots). It seems likely that sample date served as the level of replication, given that sample date was used as a random factor during comparisons between plots of different burn ages. However, sample date constitutes a repeated measure within each plot, and the analysis should have reflected this to avoid pseudoreplication during both comparisons. Thus the block structure within an ANOVA should have been specified as sample date nested within plot nested within slope position nested within catchment, the treatment burnt versus unburnt should have been specified at the catchment level, and the treatment B2, B4, B7, B15+ should have been specified at the plot level. The model could have been improved further by entering known external variables as covariates. As the analysis stands, sample sizes and hence error degrees of freedom have been badly inflated through pseudoreplication, meaning that the WTD results presented within the paper are unreliable.

Comparisons of OFLW between different burn age treatments (B2, B4, B7 and B15+ plots within Bull Clough, and U plots within Oakner Clough) were made by calculating the average proportion of crest-stage tubes indicating overland-flow occurrence per plot per visit. The authors did not use hypothesis tests to compare OFLW between different burn age treatments. Given the environmental differences between burnt and unburnt catchments and plots, it is difficult to tease apart whether differences in overland flow within the U plots are solely due to burning management. Moreover, because U plots are all located within a single catchment and all burnt plots are located in another separate catchment, the burnt/unburnt treatment is confounded with catchment, and there are no degrees of freedom for error (no statistical test is possible). Conversely, given that all the burnt plots were within the same catchment and stratified by slope, differences in overland flow between plots of different burn ages are likely to be relatively robust. However, there is a question mark about differences in slope between plots of different burn ages (See Fig.6a in Ashby & Heinemeyer 2019).

Appendix A

Extraction and calculation of Noble et al. (2017) data

We used the EMBER data uploaded by Noble *et al*. (2017a; filename: EMBER_data.xlsx) to calculate mean catchment values for the following variables: Peat depth (cm), Veg height (cm), and Bare peat (%). These response variables were measured within 12 soil plots in each catchment. The catchment values of each variable we present represent a mean of the 12 individual values (we omitted the three 'additional' plots within Great Eggleshope Beck because they do not feature in any of the EMBER papers we criticise; plots: E13, E14 and E15). We also extracted catchment values for the following variables: NH_3 deposition, N deposition, Nox deposition, O_3 deposition, SO_2 deposition, Acid deposition. These data were measured at a 5 km resolution (Noble *et al.* 2018).

References ignored by B&H during the assessment of sponsorship bias

Table A1. Burning-related references overlooked by Brown & Holden (2019) during their assessment of sponsorship bias. References were obtained using Web of Science and literature review searches (Glaves *et al*., 2013; Heinemeyer and Vallack, 2015; Harper *et al*., 2018). The search took place on the 27th of September 2019.

1	Alday, J.G., Santana, V.M., Lee, H., Allen, K.A. & Marrs, R.H. (2015) Above-ground biomass accumulation patterns in moorlands after prescribed burning and low-intensity grazing. Perspectives in Plant Ecology, Evolution and Systematics, 17, 388-396.
2	Amar, A., Grant, M., Buchanan, G., Sim, I., Wilson, J., Pearce-Higgins, J.W. & Redpath, S. (2011) Exploring the relationships between wader declines and current land-use in the British uplands. Bird Study, 58, 13-26.
3	Anderson, B.J., Arroyo, B.E., Collingham, Y.C., Etheridge, B., Fernandez-De-Simon, J., Gillings, S., Gregory, R.D., Leckie, F.M., Sim, I.M. & Thomas, C.D. (2009) Using distribution models to test alternative hypotheses about a species' environmental limits and recovery prospects. Biological Conservation, 142, 488-499.
4	Armstrong, A., Holden, J., Luxton, K. & Quinton, J. (2012) Multi-scale relationship between peatland vegetation type and dissolved organic carbon concentration. Ecological Engineering, 47, 182-188.
5	Britton, A., Helliwell, R., Fisher, J. & Gibbs, S. (2008) Interactive effects of nitrogen deposition and fire on plant and soil chemistry in an alpine heathland. Environmental Pollution, 156, 409-416.
6	Britton, A.J. & Fisher, J.M. (2008) Growth responses of low-alpine dwarf-shrub heath species to nitrogen deposition and management. Environmental Pollution, 153, 564-573.
7	Cotton, D.E. & Hale, W.H. (1994) Effectiveness of cutting as an alternative to burning in the management of *Calluna vulgaris* moorland: results of an experimental field trial. Journal of Environmental Management, 40, 155-159.
8	Dallimer, M., Skinner, A.M., Davies, Z.G., Armsworth, P.R. & Gaston, K.J. (2012) Multiple habitat associations: the role of offsite habitat in determining onsite avian density and species richness. Ecography, 35, 134-145.
9	Dennis, R.L. & Eales, H.T. (1997) Patch occupancy in *Coenonympha tullia* (Muller, 1764) (Lepidoptera: Satyrinae): habitat quality matters as much as patch size and isolation. Journal of Insect Conservation, 1, 167-176.

Table A1. Continued.

10	Dennis, R.L. & Eales, H.T. (1999) Probability of site occupancy in the large heath butterfly *Coenonympha tullia* determined from geographical and ecological data. Biological Conservation, 87, 295-301.
11	Douglas, D.J., Bellamy, P.E., Stephen, L.S., Pearce–Higgins, J.W., Wilson, J.D. & Grant, M.C. (2014) Upland land use predicts population decline in a globally near-threatened wader. Journal of Applied Ecology, 51, 194-203.
12	Douglas, D.J., Beresford, A., Selvidge, J., Garnett, S., Buchanan, G.M., Gullett, P. & Grant, M.C. (2017) Changes in upland bird abundances show associations with moorland management. Bird Study, 64, 242-254.
13	Ellis, C.J. (2008) Interactions between hydrology, burning and contrasting plant groups during the millennial-scale development of sub-montane wet heath. Journal of Vegetation Science, 19, 693-704.
14	Fyfe, R.M. & Woodbridge, J. (2012) Differences in time and space in vegetation patterning: analysis of pollen data from Dartmoor, UK. Landscape Ecology, 27, 745-760.
15	Fyfe, R.M., Ombashi, H., Davies, H.J. & Head, K. (2018) Quantified moorland vegetation and assessment of the role of burning over the past five millennia. Journal of Vegetation Science, 29, 393-403.
16	Garnett, M., Ineson, P. & Stevenson, A. (2000) Effects of burning and grazing on carbon sequestration in a Pennine blanket bog, UK. The Holocene, 10, 729-736.
17	Grayson, R., Kay, P., Foulger, M. & Gledhill, S. (2012) A GIS based MCE model for identifying water colour generation potential in UK upland drinking water supply catchments. Journal of Hydrology, 420, 37-45.
18	Hancock, M., Egan, S., Summers, R., Cowie, N., Amphlett, A., Rao, S. & Hamilton, A. (2005) The effect of experimental prescribed fire on the establishment of Scots pine *Pinus sylvestris* seedlings on heather *Calluna vulgaris* moorland. Forest Ecology and Management, 212, 199-213.
19	Hancock, M.H., Amphlett, A., Proctor, R., Dugan, D., Willi, J., Harvey, P. & Summers, R.W. (2011) Burning and mowing as habitat management for capercaillie *Tetrao urogallus*: an experimental test. Forest Ecology and Management, 262, 509-521.
20	Hancock, M.H., Summers, R.W., Amphlett, A. & Willi, J. (2009) Testing prescribed fire as a tool to promote Scots pine *Pinus sylvestris* regeneration. European Journal of Forest Research, 128, 319.
21	Harris, M., Allen, K., Le Duc, M., Eyre, G. & Marrs, R. (2011) Prescribed fire characteristics and biomass reduction on upland moorland. Aspects of Applied Biology, 171-177.
22	Heinemeyer, A., Berry, R., Sloan, T.J., 2019a. Assessing soil compaction and micro-topography impacts of alternative heather cutting as compared to burning as part of grouse moor management on blanket bog. Peerj 7, e7298.
23	Hobbs, R.J. (1984) Length of burning rotation and community composition in high-level *Calluna-Eriophorum* bog in N England. Vegetatio, 57, 129-136.
24	Lance, A.N. (1983) Performance of sheep on unburned and serially burned blanket bog in western Ireland. Journal of Applied Ecology, 767-775.
25	Lee, H., Alday, J.G., Rosenburgh, A., Harris, M., McAllister, H. & Marrs, R.H. (2013b) Change in propagule banks during prescribed burning: a tale of two contrasting moorlands. Biological Conservation, 165, 187-197.

Table A1. Continued.

26	Liepert, C., Gardner, S. & Rees, S. (1993) Managing heather moorland: impacts of burning and cutting on *Calluna* regeneration. Journal of Environmental Planning and Management, 36, 283-293.
27	Marrs, R.H., Phillips, J.D.P., Todd, P.A., Ghorbani, J. & Le Duc, M.G. (2004) Control of *Molinia caerulea* on upland moors. Journal of Applied Ecology, 41, 398-411.
28	McFerran, D.M., McAdam, J.H. & Montgomery, W.I. (1995) The impact of burning and grazing on heathland plants and invertebrates in County Antrim. Biology and Environment - Proceedings of the Royal Irish Academy, 95B, 1-17.
29	Morton, P.A., Heinemeyer, A., 2019. Bog breathing: the extent of peat shrinkage and expansion on blanket bogs in relation to water table, heather management and dominant vegetation and its implications for carbon stock assessments. Wetlands Ecology and Management 27, 467-482.
30	Newey, S., Mustin, K., Bryce, R., Fielding, D., Redpath, S., Bunnefeld, N., Daniel, B. & Irvine, R.J. (2016) Impact of management on avian communities in the Scottish Highlands. Plos One, 11, e0155473.
31	Palmer, S. & Bacon, P. (2001) The utilization of heather moorland by territorial Red Grouse *Lagopus lagopus scoticus*. Ibis, 143, 222-232.
32	Pearce-Higgins, J.W. & Grant, M.C. (2006) Relationships between bird abundance and the composition and structure of moorland vegetation. Bird Study, 53, 112-125.
33	Picozzi, N. (1968) Grouse bags in relation to the management and geology of heather moors. Journal of Applied Ecology, 5, 483-488.
34	Rawes, M. & Hobbs, R. (1979) Management of semi-natural blanket bog in the northern Pennines. The Journal of Ecology, 67, 789-807.
35	Robertson, G.S., Newborn, D., Richardson, M. & Baines, D. (2017) Does rotational heather burning increase red grouse abundance and breeding success on moors in northern England? Wildlife Biology, wlb.00276, 1-12.
36	Ross, S., Adamson, H. & Moon, A. (2003) Evaluating management techniques for controlling *Molinia caerulea* and enhancing *Calluna vulgaris* on upland wet heathland in northern England, UK. Agriculture, Ecosystems & Environment, 97, 39-49.
37	Smith, A.A., Redpath, S.M., Campbell, S.T. & Thirgood, S.J. (2001) Meadow pipits, red grouse and the habitat characteristics of managed grouse moors. Journal of Applied Ecology, 38, 390-400.
38	Tharme, A., Green, R., Baines, D., Bainbridge, I. & O'brien, M. (2001) The effect of management for red grouse shooting on the population density of breeding birds on heather-dominated moorland. Journal of Applied Ecology, 38, 439-457.
39	Todd, P.A., Phillips, J.D.P., Putwain, P.D. & Marrs, R.H. (2000) Control of *Molinia caerulea* on moorland. Grass and Forage Science, 55, 181-191.
40	Turner, T.E. & Swindles, G.T. (2012) Ecology of Testate Amoebae in Moorland with a Complex Fire History: Implications for Ecosystem Monitoring and Sustainable Land Management. Protist, 163, 844-855.
41	Usher, M.B. (1992) Management and diversity of arthropods in *Calluna* heathland. Biodiversity & Conservation, 1, 63-79.

Table A1. Continued.

42	Whitehead, S.C. & Baines, D. (2018) Moorland vegetation responses following prescribed burning on blanket peat. International Journal of Wildland Fire, 27, 658-664.
43	Whittingham, M., Percival, S. & Brown, A. (2000) Time budgets and foraging of breeding golden plover *Pluvialis apricaria*. Journal of Applied Ecology, 37, 632-646.
44	Whittingham, M.J., Percival, S.M. & Brown, A.F. (2002) Nest-site selection by golden plover: why do shorebirds avoid nesting on slopes? Journal of Avian Biology, 33, 184-190.
45	Yeo, M. & Blackstock, T. (2002) A vegetation analysis of the pastoral landscapes of upland Wales, UK. Journal of Vegetation Science, 13, 803-816.