Exhibit A: Studies Cited in the Buffer Report Fail DQA Standards

<u>Johnson *et al.* 2011</u>

This paper seeks to determine whether specific activities are correlated with population level declines in sage grouse, as determined from lek count trend data. The idea is to identify quantifiable threats to sage grouse populations.

The authors examined 62 different variables (Table 1) using only 11 years of lek count data for the response variable in seven different management zones. This study is an example of an extremely weak approach to statistical inference and a poorly planned data-fishing expedition. There are simply not enough years of data to support inferences with single variables, much less several variables. By chance alone, several variables should show correlations with lek count trends. The problem is compounded by the fact that many of the lek counts had *only four years of data* associated with them.

From the "Conservation Implications" section at the end of this paper you would not know that lek counts have generally increased over the 10-year period that this study looked at (Figure 2), although the authors have several convenient caveats to explain this away.

Basically, the figures tell the story, that there are no significant correlations between predictor and response variables. These are essentially random clouds of points. The authors resort to loess smoothing to search for patterns in the data that do not have obvious statistical significance. Despite this, the authors report on "trends" and discuss the potential importance of these in the paper.

Consequently, the resolution of the data and the methods applied to them is extremely limited. The authors admission of limitations and caveats is not enough to salvage the results or redeem weak inferences based on them. Had this paper undergone a rigorous and independent peer-review, it would have almost certainly been rejected. It is doubtful that this paper would be considered publishable in most reputable scientific journals.

The authors used data from 9,844 leks but *"only the 3,679 leks with at least four annual counts during the 11-yr period were included* [in analyses]." In comparison, Garton et al. (Chapter 16 used data from 9,780 leks. The reason for the difference between two studies using the same data, in the same monograph, with many of the same coauthors is not explained.

In the last lines of the paper, the authors voice a number platitudes that are consistent with the message of other papers in this monograph: *"No single factor is responsible for declining sage-grouse populations, and no single action may be sufficient to restore them. Conservation of the species will initially require a recognition of the intrinsic value of sagebrush-dominated landscapes, followed by the development of a comprehensive approach to sagebrush habitat conservation that involves commitments and partnerships among state and federal agencies, academia, industry, private organizations, and landowners; Knick et al. (2003:627) affirm that* *only through this concerted effort and commitment can we afford to be optimistic about the future of sagebrush ecosystems and their avifauna."*

Blickley *et al.* 2012

This study reported a population decline in lek attendance when projected sound from recordings at the edges of leks, which were as high as the noise levels occurring within 200m of a busy freeway (as measured across an open field with traffic loads of greater than 50,000 cars per day, or 55-70 decibels as shown in Figure 2 of Reijnen et al. 1995). The subsequent avoidance was then assumed to lead to have a negative effect on the population (i.e. contribute to their decline). Below, is a relevant excerpt from Blickley et al.:

> Drilling-noise recordings were broadcast on experimental leks at an equivalent sound level (L_{eq}) of 71.4±1.7 dBF (unweighted decibels) SPL re 20 μPa (56.1±0.5 dBA [A-weighted decibels]) as measured at 16 meters; on road-noise leks, where the amplitude of the noise varied with the simulated passing of vehicles, noise was broadcast at an L_{max} (maximum RMS amplitude) of 67.6±2.0 dBF SPL (51.7±0.8 dBA).

The fact that authors broadcast such high levels of noise in such close proximity to leks biased the results, an error of omission by the authors and the Buffer Report that cites them and proposed regulations based upon their recommendations.

The Buffer Report cannot have it both ways, claiming a negative effect on sage grouse populations but admitting that there was "low statistical support for a cumulative effect of noise over time" in the study by Blickley *et al*. As noted above, there are no data showing a long-term cumulative decline in the sage grouse population in the Pinedale Planning Area.

The cited research was an amateurish attempt to reproduce the sounds of oil and gas development using substandard equipment that was wholly unsuited to the task of accurately recording and playing back traffic and sounds from oil and gas operations. Deficiencies in Blickley et al.'s equipment are detailed below.

Microphone: According to the manufacturer (http://en-us.sennheiser.com/k6-microphone-system), "the ME 62 [microphone used by Blickley et al.] is an omni-directional microphone head suitable for K6 and K6P powering modules. It can be used for reporting, discussions and interviews. The ME 62 is particularly suitable for good reproduction of 'room' ambience and 'spaced omni' stereo recording. Matt black, anodized, scratch-resistant finish."

Recorder: The Marantz model PMD670 used by Blickley et al. does not offer high-resolution (88.2 or 96 KS/s) sampling rates, its metering characteristics are unknown, and it is limited to 16/48 recording and thus is not considered a high-resolution recorder. It retails online for $700.

Playback speakers: The speakers used in the study were standard outdoor speakers camouflaged as rocks and designed for background music playing in home, hotel, and amusement park applications. They were not designed for accurately reproducing industrial sounds. The specifications for the speakers may be found on the manufacturer’s website:
http://www.ticcorp.com/specifications_tfs14.pdf.

The speakers were powered by 12 volt car batteries rather than 120 volt AC power and a car stereo amplifier of unknown make and model was used to boost the output. Packed into each simulated rock speaker housing was a 10" woofer with an injection molded cone, a 5.5" midrange cone, and 2" soft dome tweeter. The size and quality of the speakers, and the small speaker housing, severely limits the physical capability of the system to accurately reproduce either low or high frequency sound produced by oil and gas operations or traffic.

As a result of substandard equipment and lack of expertise in sound recording and reproduction, Blickley *et al.* resorted to placing their speakers at the edge of leks and to playing their systems at high levels in order to elicit a behavioral response. This is a biased approach to obtain a preferred result. The BLM cannot rely on biased research in its decision-making.

The recommended noise levels are not based upon any standardized, repeatable data collection, or accepted methods of sound measurement. The methods used by Blickley *et al.* and reported results did not contain any credible, professional analysis of local ambient sound levels or oil and gas noise (e.g. the type, duration, frequencies, sound pressure levels, and power of sound produced by different oil and gas drilling or production operations; equipment being recorded); or employ the use of professionally accepted standards, such as International Organization for Standardization (ISO) standards for quantifying industrial and traffic noise (http://www.iso.org/iso/home/standards.htm).

The standards not followed by the cited studies include, but are not limited to: ISO 1996-1:2003 Acoustics -- Description, measurement and assessment of environmental noise -- Part 1: Basic quantities and assessment procedures; ISO 9613-2:1996 Acoustics -- Attenuation of sound during propagation outdoors -- Part 2: General method of calculation; ISO 4871:1996 Acoustics -- Declaration and verification of noise emission values of machinery and equipment; ISO 532:1975 Acoustics -- Method for calculating loudness level; ISO 7196:1995. Acoustics -- Frequency-weighting characteristic for infrasound measurements; ISO 8297:1994 Acoustics -- Determination of sound power levels of multisource industrial plants for evaluation of sound pressure levels in the environment -- Engineering method; and IEC 61672-1:2002(E) - Electroacoustics, Sound level meters -- Part 1: Specifications).

Blickley *et al.* did not employ any sound propagation models in their study to quantify the confounding effect of temperature, relative humidity, topography, ground cover and surface porosity, wind direction, the direction noise was generated from, the geographic extent of the noise, its duration, frequency of occurrence, or permanence, (Attenborough 2007). Nor did they provide any correlation of their playbacks compared to the industrial and traffic sources they had attempted to duplicate. Furthermore, no graphic equalizer was used which would have allowed for the adjustment of sound pressures in different frequency ranges (at standardized 1/3 octave band frequencies), and no measurement of sound pressure levels was taken in front of playback speakers, which together would have allowed for the accurate reproduction of the sound at the same frequencies and sound pressure levels as the original noise. Therefore, BLM cannot base regulations upon no data and results based upon arbitrary methods that are not compliant with accepted professional standards in the noise control industry (i.e. Bies and Hansen 2009; ISO).

Holloran and Anderson 2005

1) Management recommendations of Holloran and Anderson (2005) are opinions.

This paper use the results of statistical tests involving data gathered on sage grouse lek-to-nest distances, nest-to-nest distances, and nest success, as a bully pulpit to justify a number of sweeping management recommendations that are unrelated to the few statistically significant results. The authors appear to have a preconceived notion that a five km buffer surrounding leks is needed to "protect" sage grouse nests, however, their study did not quantify any anthropogenic threats or explain why the proposed buffer would protect them. Moreover, according to the authors, the study was carried out in areas "free of large scale habitat conversions" and "areas fragmented by oil and gas development were removed from consideration." Therefore, the recommendations made concerning buffers from human disturbance/activity are nothing more than unsupported opinions.

2) The selection of 5 km distance threshold for buffers is arbitrary.
It was simply an arbitrary threshold, in the authors' opinion, that "suggested that a 5 km buffer around a lek was needed to encompass a relative majority of nests" (in this case 64% of nests). And as noted above, all those leks and nests were in areas chosen to be unaffected by human disturbance and habitat alterations. No study was conducted to understand what buffers should be in something other than a nearly pristine landscape of continuous sagebrush, or what effect buffers or a lack of them would have on a population as a whole. Furthermore, buffers do not address any specific cause and effect mechanism associated with identified threats.

3) Data used in analyses by Holloran and Anderson 2005 are not public, rendering their results irreproducible.
The data used in Holloran and Anderson (2005) are not public so their results are not reproducible. Additionally, Holloran and Anderson (2005), and Holloran (2005, that they refer to for additional detail), did not identify any of the leks by name or identifier that could be used to trace their location through the State of Wyoming's sage grouse database. Moreover, the only source of information on leks, Figure 1 of Holloran (2005), only portrays 21 leks on a map at low resolution. It is never explained how both Holloran (2005) and Holloran and Anderson (2005) claim that female sage grouse were captured from "30 relatively undisturbed leks throughout central and western Wyoming" but not provide any further information on the name and approximate location of leks. One can only wonder where the other 9 study leks were located. Thus, the location of leks where females were captured, when they were captured, the habitat they were captured and nested in, the proximity to other leks, and sage grouse density, are all undisclosed precluding any replication of results. These facts render the results of Holloran and Anderson (2005) as irreproducible.

4) Holloran and Anderson (2005) use linguistics to make statistically insignificant results sound as if they represented actual trends rather than negative results.
Statistical inference is simple and straightforward: either the result is statistically significant or it isn’t. Holloran and Anderson (2005) however, presented results in the discussion in a way that suggest these represented trends, despite their obvious lack of statistical significance. A lack of statistical significance did not appear to stop these researchers from finding the answer they were looking for.

For example, in the abstract of Holloran and Anderson (2005), two non-significant results were presented as if they represented a biologically significant tendency.

"Closest known lek-to-nest distance was greater for successfully hatched compared to destroyed nests, and closely spaced nests tended to experience lower success and have higher probabilities of both nests experiencing the same fate compared to isolated nests, suggesting that a mechanism of enhanced prey detection occurred at higher nest densities."

However, those reported results were not statistically significant, and therefore, do not suggest anything of the sort. The only potential explanations are: no relationship or a lack of resolution in the data, both of which are presently indistinguishable.

From Holloran and Anderson (2005) Results (emphasis added in bold):

"However, nests located <1 km from another known nest (n = 58) ***tended to have lower than expected probability of success*** *(cumulative 28%; X2 = 3.5; p = 0.06), and the probability of both nests (n = 38 pairs) experiencing the same fate (cumulative 71%)* ***tended to be greater than expected by chance*** *(X2 = 3.0; p = 0.08)."*

From Holloran and Anderson (2005) Discussion:

Additionally, nests located within 1 km of another known nest **tended to have lower success probabilities**, suggesting that increased nest densities could negatively influence the probability of a successful hatch.

The minimum threshold for statistical significance typically used in statistical inference is $p < 0.05$, and this was also the level of statistical significance used by Holloran and Anderson (2005): *"statistical significance was assumed at p < 0.05."* No mention was made by those authors of any alternative interpretations of non-significant results. Inexplicably, however, the results from p values of statistical tests exceeding that threshold, (i.e. p = 0.06 and 0.08), were referred to as *"tended."* It is difficult to describe this as something other than a misreporting or misrepresentation of results. Results that were clearly not statistically significant should have been reported as something else.

5) Non-significant results were used in Holloran and Anderson (2005) to "support" hypotheses: a practice that places their conclusions outside the realm of science.

Perhaps more disturbing than the example above, was Holloran and Anderson's (2005) use of those same insignificant results to "support" various hypotheses and suggest several biologically-significant phenomena. Examples are provided below:

"Our results suggest that Greater Sage-Grouse nests located relatively near (within 1 km) another known nest ***tended*** *to be less likely to successfully hatch,* ***supporting*** *this hypothesis."*

"Pairs of nests spaced relatively closely within 8.5 km of a lek ***tended*** *to experience the same fate more frequently than was expected by chance,* ***suggesting*** *predators concentrated search effort in specific areas (Niemuth and Boyce 1995), and* ***supporting*** *the idea of behavioral changes by predators. Our results* ***suggest*** *that a mechanism of enhanced prey detection occurs at higher nest concentrations, and that increased nest densities could result in increased nest depredation."*

In both cases, the results were not statistically significant, therefore, the results do not have any tendency that can be distinguished from chance alone.

5) Multiple comparisons not corrected for, thus invalidating the reported statistical significance.

In addition to the issue above is the fact that Holloran and Anderson (2005) did not employ even the most basic statistical procedure to correct their threshold of statistical significance based on the number of comparisons (17 comparisons with tests of significance and 6 correlations without tests of significance), and many of these involved multiple uses of the same variable across multiple tests. Minimally, if Holloran and Anderson (2005) had adjusted for statistical significance, at least two of their five reported significant comparisons would have evaporated, as they were at or close to the $p<0.05$ significance threshold.

The reason to perform corrections for multiple comparisons, (using procedures such as the Bonferroni or Bonferroni-Holms corrections (Holm 1979) in the case of independent tests and the Benjamini–Hochberg–Yekutieli procedure (Benjamini and Yekutieli 2001) in the case of dependent tests where all or a portion of data are shared across multiple tests), is to minimize the Type-1 error or false discovery rate associated with performing multiple tests. This minimizes the chance of reporting erroneous statistical significance that would be expected to increase as more tests are performed). This is done by adjusting the threshold value of significance to be more stringent, thus minimizing the chance of reporting erroneous statistical significance that was the result of chance alone.

In the case of Holloran and Anderson (2005), applying the Bonferroni procedure would have rendered all of their the p-values non-significant, and if only minimally applied to cases of shared data (e.g. lek to nest distances), only two tests would remain significant: the number of nests vs. distance from lek in all 0.5 km bands within 3 km, and distance to next years nest after unsuccessful vs. successful. Those results would leave virtually little or nothing interesting for Holloran and Anderson (2005) to write about, and further refutes their extensive discussion of a 5 km buffer and similar management recommendations.