

I appreciate the opportunity to review this paper, which presents an interesting and important possibility that grizzly bear fatalities caused by non-hunting humans is under-reported. I believe that the essential conclusion of the paper is likely correct, but I also believe that the evidence leading to the conclusion is not yet entirely convincing. The approach of using a training data set of grizzly bear fatalities to inform the estimate of non-hunter killed bears in the larger population has real potential, and raises interesting questions and challenges related to ensuring the estimate is accurate.

Grizzly bears participating in a telemetry study that was begun in 1979 are used in this paper as a training data set to obtain what are assumed to be the true proportions of bears killed by hunters versus non-hunters. The ratio of hunter to non-hunter fatalities in the training data is then used to adjust an agency data base of bear fatalities for the proportion of non-hunter deaths that were unreported. This is an approach often used to adjust fatality estimates for the proportion of fatalities not found during fatality monitoring where human activities caused wildlife deaths. The number of detected fatalities F is divided by a fatality detection rate D informed by trials imposed on the searchers who, ideally, are blind to the trials. The essential form of the adjustment, $\hat{F} = \frac{F}{D}$, has its origin in Horvitz and Thompson (1952), but multiple revisions or additions have since been made to their estimator to suit specific challenges (Korner-Nievergelt et al. 2011). Applied to the bear study, F is the number of bears reported killed by hunters, D is the ratio of hunter to non-hunter killed bears in the training data, i.e., the telemetered bears, and $\hat{F}$ is the estimated number of bears killed by humans who were not hunting.

Before continuing, I should interject three points. One is that the paper needs to simplify the analysis by eliminating comparisons of outcomes by gender. The case for gender differences in reported fatality rates is weak, where percentage differences really represent small numerical differences in bears of one gender or the other. Second, whether to include suspected causes of death should be decided, and either the results should include or exclude fatalities of suspected cause. I suggest using only bears of known cause of death, but I doubt that it makes much difference either way. Third, the study results should include only one time period, which is the period over which bears were captured for the telemetry study. Breaking out results over a more recent time period is justified by an assumed greater reporting accuracy in recent years, but no evidence is provided in support of this assumption. Why would reporting be less accurate during 1980-1995 than during 1995-2015? Using the more recent time period yields a greater adjustment to the number of bears killed by non-hunting humans, but at the cost of relying on an even smaller sample size. I suggest using only the study period 1980-2015. In summary, the paper would be stronger by restricting its focus to those bears reported as fatalities caused by hunters and non-hunters and with and without telemetry units, i.e., only four numbers. These four numbers would be 10 collared bears killed by hunters, 9 collared bears killed by non-hunters, 70 uncollared bears killed by hunters, and 10 uncollared bears killed by non-hunters.

The Horvitz-Thompson estimator, or any of its analytical descendants, can be highly sensitive to the effects of bias and error. A relatively small change in D can greatly affect the adjusted number of fatalities. If only one fatality was found or reported, whereas 50% of the training data were found or reported, then $1 \div 0.5$ yields 2. But if $D = 0.1$, then the adjusted fatalities increases to 10 ($1 \div 0.1$), and if $D = 0.01$, then the adjusted fatalities increases to 100 ($1 \div 0.01$). The outcome of the fatality reporting of telemetered bears makes a huge difference to the

estimated number of bears killed by non-hunters. Therefore, the training data set used to derive D must be reliable.

The reliability of the training data in the bear study *bears* scrutiny for several reasons. D is a ratio unaccompanied by an error term. The ratio that was reported for the training data was 1.11 ($10 \div 9$), which would yield an adjusted number of non-hunter fatalities of 63. However, D was based on a small sample size, meaning that a shift of one fatality from the hunter to non-hunter category, or vice versa, would change D to either 0.90 or 1.375 depending on the direction of the shift. Applying these outcomes to 70 reported hunter deaths, the adjusted non-hunter fatalities could be either 51 or 78 bears killed by non-hunting humans, or 19% lower or 25% higher than the reported number of 63. A shift of only 2 fatalities to the other category would change D to 0.73 or 1.71 depending on the direction of the shift. Applying these outcomes to 70 reported hunter deaths, the adjusted non-hunter fatalities could be either 41 or 96 bears killed by non-hunting humans, or 35% lower or 52% higher than the reported estimate of 63. The paper should include an assessment of uncertainty of the study result due to small sample size.

Another reason to scrutinize the reliability of the training data goes to the assumption that the ratio of hunter to non-hunter fatalities of telemetered bears represents the same ratio applied to non-telemetered bears. The paper implies that this assumption is valid because the telemetry units were small and difficult to see from a distance (line 107). However, this assumption could be flawed for two reasons that were not discussed in the paper. First, hunters using scoped rifles were likely to notice the telemetry units. If only one bear was spared by a hunter who noticed the telemetry, and had that bear later died as a result of an encounter with a non-hunter (train, car, or self-defense), then the adjusted fatalities due to non-hunting human causes would have shifted from 51 to 63. Had two hunters made this decision and both bears later died due to non-hunting human causes, the shift would have been from 41 to 63. Secondly, the assumption might be flawed if telemetered bears shift their behaviors as a result of capture and handling. Alternatively, telemetered bears might have been captured because their behaviors differed from non-captured bears, making them easier to capture. For fatality rates of telemetered bears to represent those of non-telemetered bears, the telemetered bears would have had to have been a random sample from the bear population and their behavior unaltered as a result of capture and handling.

Lack of experimental design adds another reason for interpreting the result cautiously. Even a mensurative study can benefit from the basic tenets of experimental design such as replication and interspersion of treatments, use of a control treatment, and appropriate spatial and temporal scales (Hurlbert 1984). As a wildlife ecologist I understand that studies of species such as grizzly bear rarely allow for the implementation of experimental design tenets. In the case of this study, the treatments of hunters and non-hunter killers of bears were mixed but not interspersed in the same study area, and there was no replication and no control treatment (sizable areas with no hunting allowed). Although experimental design tenets are understandably absent or at best weak due to the nature of the animal, the study result should be interpreted cautiously. I suggest that the paper present the under-reporting of non-hunter-killed bears as a possibility – a possibility that warrants focused research on the question.

Even in raising the possibility of under-reporting of non-hunter killed bears, I suggest that it would help to add more details about the bears reported killed this way. The paper lists the types of causes of death other than hunting, but because the under-reporting of non-hunter deaths is central to the paper, I suggest adding a table that summarizes the circumstances associated with each bear assigned to this category. How many were killed in self-defense? How many were killed by cars? It would help to report whether any of the bears assigned to the non-hunter fatality category carried bullets or showed other evidence of wounding by hunters. Confounding factors affecting only one or two bears in this study can greatly change the study's outcome.

The reporting would benefit from some changes. The terminology could use some work. For example, 'mortality' is often used where 'fatality' would be more appropriate, as fatality refers to a death event whereas mortality refers to a rate, e.g., deaths per 100,000. The term 'mortality rate' is redundant.

'The number of bears killed by people for all other reasons [than hunting]' might be more concisely represented by 'the number of non-hunter-killed bears.' Basically the study is about bears killed either by hunters or non-hunters.

Specific editing comments

Line 3 I suggest breaking the first sentence into two sentences, the first ending with conservation. I suggest merging the next sentence by replacing the period with a comma and the 'however' with a 'but'.

Line 8 Replace 'for' with 'including'.

Line 9 Delete the first 'for'.

Line 15 Add comma after 'office'.

Line 20 Delete hanging parenthesis.

Line 39 The reasons for non-hunter fatalities ought to be summarized in the Introduction.

Line 114 Table 1 identifies the number of female bears as 37. However, I suggest skipping all of the analysis of gender differences in reporting of bear fatalities.

Line 118 Here is a good example of why it would help to include a table summarizing the circumstances of fatalities assigned to the non-hunter category. This bear's collar was attached to bottles and tossed into the river, probably bringing some laughs to whoever did it. But how does this act of research vandalism, by itself, support the determination that the bear was killed by the non-hunting public? What if the vandal(s) found the bear dead already, either killed by a hunter or by natural causes? Left as is, the assignment of the bear to the non-hunter fatality category seems like a leap to guilt. There must be an additional reason for the category assignment.

Line 122 This entire paragraph and much of the next paragraph can be deleted without loss to the main result of the study.

Line 141 This paragraph is discussion material, so belongs in the Discussion section. I also suggest deleting all discussion about sex differences in reporting rates.

Line 164 Yes, the sample size is small. For this reason, I suggest revising the following clause by adding a statement of uncertainty such as ‘might indicate.’

Line 169 Replace included with include.

Line 174 The second reason seems the same as the first.

Line 175 The sentence beginning with ‘In addition’ lost me. I suggest rewriting it.

Line 179 Is there any basis for this speculation about more male bears being killed in hunter camps?

Line 192 Why would a hunter report not shooting a bear because it was collared?

Line 194 This paragraph includes multiple conclusions that are unfounded or overconfident. Some of the statements could use citations, and some could use more caution.

Line 204 The citation does not appear in the list of references.

Line 207 I suggest deleting the sentence on the time period beginning in 1995.

Line 208 I could not understand the sentence about managers and researchers knowing something by far...

Line 211 ‘...nature of human-caused mortalities...’ seems vague. It would help to be more specific.

Check the referencing. McLellan 1998 is cited on line 21, but is not listed in references. The same for Servheen. Check all of them.

References Cited

Horvitz, D. G. and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association* 47:663–685.

Hurlbert S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54:187-211.

Korner-Nievergelt, F., P. Korner-Nievergelt, O. Behr, I. Niermann, R. Brinkmann, and B. Hellriegel. 2011. A new method to determine bird and bat fatality at wind energy turbines from carcass searches. *Wildlife Biology* 17:350–363.