

**SRC COMMENTS ON AUGUST 2008 FATALITY MONITORING REPORT, M21**

2 September 2008

Alameda County Scientific Review Committee

The Altamont Pass Avian Monitoring Team recently produced a draft monitoring report of fatalities and utilization rates in the Altamont Pass Wind Resource Area (APWRA). The Scientific Review Committee (SRC) reviewed the report and discussed its comments with the Monitoring Team during a conference call on 14 August 2008. Because the Monitoring Team expressed reluctance to immediately revise the report according to the SRC's comments, the SRC agreed to submit this statement of concerns to Alameda County. An overview of the SRC's concerns appears on this page, and detailed comments on the report appear in the following pages.

Whereas the SRC found multiple reasons to praise the Monitoring Team for its latest draft of M21, it also expressed frustration with the following patterns:

- Multiple SRC recommendations were not implemented;
- The Monitoring Team did not involve the SRC Subcommittee that had been formed to help the Monitoring Team develop the report so that it is consistent with SRC recommendations;
- Conclusions of mortality changes were presented with more certainty than the SRC felt was justified;
- Limitations of the data and methodology were not discussed.

The SRC is skeptical of the mortality estimates and changes in mortality estimates between the baseline and current monitoring programs. The major methodological reasons for the SRC's skepticism are the following:

- The Monitoring Team used the midpoint value in ranges of estimated days since death of bird carcasses, but these ranges were large in most cases (0-90 days, 31-90 days, etc.) and the implied error was not carried through the calculation of mortality;
- The Monitoring Team excluded 36% of the fatality records gathered in the baseline effort, but excluded only 17% of the records from the more recent effort, thus possibly biasing mortality higher in the more recent effort;
- The Monitoring Team excluded fatalities that occurred prior to the first fatality search at each wind turbine, thereby using only one search per turbine at 2,548 turbines when estimating baseline mortality;

- The analysis and presentation of the bird utilization data were confusing, and possibly biased due to the metric consisting of the number of minutes of activity rather than an index of the number of birds observed during point counts;
- The report is unclear in its definition of different metrics, such as mortality and fatality rates; and,
- Comparing fatality rates that were estimated by month produced potential monthly biases because different subsets of wind turbines were surveyed each month, so one portion of the APWRA might have been represented one month, and another portion of the APWRA was represented the next month (The SRC had recommended that the minimum time unit be season, but if the Monitoring Team was to use month anyway, then the SRC recommended a moving average to help smooth over the month-to-month differences in turbines searched).

The SRC recommends that the Monitoring Team revise M-21. If M-21 is not revised, then the SRC recommends that these comments in P-107 accompany M-21 as an attached document.

**Smallwood Review of Monitoring Team Report of July 2008 (M21)**

12 August 2008

I appreciate the monitoring team producing the report, though I think it should have been delayed longer. I think the data base needed more work and discrepancies between the monitoring team's report and my reports should have been examined carefully. I think more of the SRC's recommendations should have been followed. Also, I cannot accept certain methodological changes to the estimation of mortality, though their use did demonstrate a general robustness of the mortality estimates. Anyhow, I hope my comments are taken in the spirit of constructive scientific review, and I hope there is an opportunity to revise the report.

**Page Para. Line**

2-3 1 8

Smallwood and Thelander (2004) is an inappropriate reference here, since they neither recommended a winter shutdown or relocations of Tier 1, 2, or 3 turbines. Also, it should be clarified that Smallwood and Spiegel (2005a,b,c) recommend a 4-month shutdown, not a 2month shutdown.

Also, the paragraph addresses the settlement agreement, but says nothing about non-settling companies and their permit conditions.

2-3 2

The monitoring process was established a year and a half before the settlement agreement. This summary could be more accurate on the history of the permit conditions and the goals and objectives of the Avian Protection Program.

3-3 2 last

Probably a better citation would be of Smallwood and Thelander (2008), since it was more recent and went through peer review at a scientific journal.

3-4 1

I found the description of the point count methods a little confusing. I realize the protocol changed at least once, if not more than once, while the SRC and monitoring team deliberated this method over the past two years. Maybe this is why I'm confused by this description. The problem is that point counts typically do not involve on-the-minute recordings of bird observations. Instead, they count all the birds seen during the time period of the point count. If the point count was limited to on-the-minute observations, then all the birds that were detected between the on-the-minute recordings would not be recorded. Perhaps the monitoring team is performing point counts in the traditional manner now, but used the on-the-minute observations over the last two years to be comparable? Maybe a little clarification is needed here.

3-4 2

The third bullet is a little misleading. The search area was smaller than 125 m.

3-7 3

Why did the monitoring team use a different protocol for including/excluding fatality records found during initial surveys? The methods state that the protocol differed, but no reason for the difference was given. Does the monitoring team realize what this protocol means for the 2,548 wind turbines that were searched only twice during the Smallwood and Thelander study? Their standard leaves only one search per each of these turbines. In my opinion, this protocol makes unsuitable any APWRA-wide comparison between the baseline and recent monitoring data, unless the monitoring team omits the Set 2 turbines. Basing mortality estimates on one search is unacceptable.

3-8 Fig. 6

This figure does not inform the reader how many carcasses were backdated 0-90 days, 31-90 days and so on. I think the reader needs to know this, because most of the records were assigned large ranges of days since death. Also, these estimates are later used to estimate mortality, which I think is another fatal flaw of the report.

3-10 1

I disagree with the method of using estimates of days since death of carcasses as the adjustment term for estimating mortality. I think this is a fatal flaw of this report. The data are unsuitable for this purpose, consisting of large ranges of days since death. They were also categorically applied, leading to potential biases in estimates of days since death. I found, for example, that a disproportionate number of eagles were estimated to have died within 3 days of the search, whereas a disproportionate number of burrowing owls were estimated to have died 0-90 days since the search.

When I first heard of this method, it interested me and I thought there was some merit in investigating it. However, after reviewing the monitoring team's data, I realized that the estimates of days since death were usually very large ranges, and were unsuitable for this approach. I produced two reports (P-101 and P-97) warning that such an approach would not be appropriate.

Whereas the monitoring team routinely estimated days since death in ranges of values – usually large ranges – many of the “baseline” fatalities were reported as point values, without ranges. The method of reporting days since death shifted between studies, largely due to differences in opinions between the search crews. I'm concerned about a serious bias in this comparison.

Furthermore, it looks to me as though the error terms associated with scavenger removal and searcher detection rates were not carried through the calculations of mortality, such as by using

the Delta Method. Why not? How did the monitoring team handle the large error in the estimates of days since death? It looks to me like the team ignored this error, but I don't see how anyone could ignore error on the order of 0-90 days since death.

The methods of Smallwood (2007) and Smallwood and Thelander (2008) were peer-reviewed at the world's leading scientific journal on wildlife biology, so I'm wondering why the monitoring team abandoned use of the average search interval and in its place adopted an approach for which the data are unsuitable. I had worked out a tool-kit, so to speak, for using average search interval, which also allowed me to largely ignore the error around estimates of days since death, because days since death were not directly used to make the mortality estimates. On the other hand, the monitoring team used these shaky estimates of days since death and didn't appear to develop any means for handling the error in the estimates. I think the new approach was a good idea, but turned out to be inappropriate because the data were unsuitable for it.

3-10 last

Also, the estimation of mortality by month was inconsistent with the standing SRC recommendation (see meeting notes from February 2008), asking that the minimum units be mortality per string and per *season*.

3-12 2

The comparison of relative abundance is actually of relative activity level. Relative abundance, as usually quantified in point counts, would be of the average number of individuals detected, and not the average number of minutes the birds were observed. A perched bird would be assigned a value of 10 for 10 minutes of perching during a 10-minute point count, even though it was one bird, so one can get a 10-fold swing in "relative abundance" per record, depending on whether the number of minutes or the counts of individuals are compared. Another significant problem with this approach is that perching behavior could vary seasonally, so one could get a lot of values of 10 for golden eagle during one season and only values of 1 during another season when it really might be the same eagles perching during seasons of slower wind and flying by fleetingly during seasons of strong winds. So during one season you could have a lot of eagle presence due to perching when in fact it's the same eagles expressing different behaviors. I think the monitoring team needs to compare the average number of birds, as I recall suggesting in our last in-person meeting.

Methods Section

Nowhere in the Methods section were statistical tests described. This was an important omission because I couldn't understand the test results that appeared in the Results section.

4-1 1 6

Why was the capacity of the core turbine capacity only 50 MW? I had it at 82.63 MW (see P-76).

4-1 1 7

The baseline study involved 418.255 MW of capacity, not 396 MW. How did the monitoring team come up with 296 MW? Check Smallwood and Thelander (2004, 2008).

4-2 Table 1

The data I received from the monitoring team indicated the capacity of the regularly searched turbines other than Diablo Winds totaled 285.34 MW, but Table 1 identified 273 MW as the capacity. Why the difference?

Also, how did the monitoring team handle data from the turbines on property held by East Bay Regional Parks District? How was Buena Vista dealt with?

4-2 Fig. 8

How does MW searched per month help the reader understand the mortality estimates? Speaking for myself, I don't see the point of listing these values in a table.

4-3 Table 2

The monitoring team excluded 36% of the fatality records gathered in the baseline effort, but excluded only 17% of the records from the more recent effort. One reason for this disparity in records omissions was a difference in the number of "complete records." I'm not sure what this means. I should have asked what this means in the last in-person SRC meeting, when it was presented in outline form. My worry is that this omission standard was arbitrary and second-guesses the professional judgment exercised by the author of the baseline mortality estimates (OK, that would be me). Furthermore, I worry that this disparity in omission rates resulted in erroneous conclusions that mortality increased for some species. Furthermore, these omission rates are high – higher than I used, and I wonder whether they are appropriate. I cannot imagine tossing out more than a third of the fatality records we bothered to collect from 1998 to 2003.

4-4 Table 4

Does this table include all fatalities found, or only those that were passed by the monitoring team's filters? I ask because Table 4 states that only 26 golden eagles were found in the baseline study, but I recall 54 golden eagles were found. Maybe the table needs clarification.

4-5 Table 4

Why was red-tailed hawk "continued?" Counts of red-tailed hawk and golden eagle appeared twice in the table. Also, red-winged blackbirds are listed twice, once as a non-raptor and once as a raptor. Table 4 needs to be cleaned up.

4-6 1

The report states, “However, for all four target species, study period was a poor predictor of monthly fatality rates ( $R^2 < 0.10$  all cases).” What is this  $R^2$  value based on? What test? What is the “period?” If it is regression, isn’t the notation wrong? And how could a regression be performed on two data points? I just don’t get it.

4-6 2

The results reported by the monitoring team differ from mine using the same comparisons of time periods and turbines (see P-76), but the monitoring team’s capacity figures don’t compare to mine, and neither do their fatality record omissions and handling of scavenger removal.

4-6 2 last

Again, what are these  $R^2$  values referring to, and what is “study period?”

4-7 1

According to the report, some of the differences in mortality between Diablo and non- Diablo turbines were due to turbine type, and there’s an  $R^2$  and P-value associated with this conclusion. All the turbines in the Diablo group are of a different model than all the turbines in the comparison group, so I guess I don’t understand why the monitoring team feels it was necessary to point out that some of the variance was explained by turbine type. Am I missing something here?

4-7 1 last

The monitoring team reported that no golden eagles were killed by Diablo winds, which highlights a problem I have with their data omission standards. According to the data supplied to me by the monitoring team, Diablo Winds killed a golden eagle on about 27 August 2006, which was two months before the eagle was found by the monitoring team near a wind turbine. The carcass was complete and met all the inclusion standards (as far as I could tell) except the one that dumped all records estimated to have been caused prior to the comparison period. My method would have included this eagle, and I think it should have been included. What is the point of performing fatality searches if one is going to omit the fatalities that are found?

The way I originally dealt with the fatalities estimated to have been killed within 90 days of the first search was to include the fatalities and add 90 days to the monitoring period, so a 2 year period might become a 2.25 year period. Later, I used the average search interval for the particular group of turbines as my standard, so if a fatality was backdated before the first search, but within the average search interval, then I would include the record and I added one search interval to the monitoring period. I recall the SRC deliberating the first search problem, and it agreed that my general approach would be used. Again, the monitoring team’s new approach also reduces the number of searches among Set 2 turbines in the baseline monitoring period to 1,

which is unacceptable for estimating mortality. Using only 2 searches was bad enough (that would be my doing), but deriving mortality estimates from 1 survey won't work.

4-11 Fig. 9

Does this Figure depict numbers of minutes or numbers of birds?

4-12 Figs. 10 and 11

What became of using moving averages, as the SRC recommended?

4-12 Fig. 10b

There is no point in showing the reader this graph of a non-significant regression. It uses a lot of report space.

4-13 Fig. 11b

I would like to know what units are used in the figure. Are these minutes of activity or numbers of birds? If the former, I'd be cautious about drawing any conclusions from the relationship.

4-14 Fig. 12b

As I pointed out on July 8, this regression is inappropriate. It is dominated by one data point, the removal of which would kill the  $r^2$  value. I think this test is over-reaching. I recommend that it be dropped.

4-15 Fig. 13b

Again, there is no point in showing the reader a graph of a non-significant regression.

5-1 1

According to the report, "Analysis of the Baseline and Current data revealed marked increases in the annual mortality rates and total fatalities of most target species and species groups examined, except the golden eagle, for which fatalities decreased by 35%. When the results of all target raptor species were combined, the mortality rate for the Current Study group increased 74% over Baseline." However, I am not convinced the report leads to these conclusions. I don't believe it was appropriate to adjust mortality by the estimates of days since death, and I see significant discrepancies in the numbers of fatalities and MW compared between the reporting by the monitoring team and by me (P-76). I do not believe the mortality of target species increased, let alone increased by 76%.

A-1 2 1

This fatality definition has been challenged by the SRC since the SRC first saw it. As late as 11 July 2007 the SRC provided written comments challenging this definition (also see P-39).

C-2

The source of these scavenger probabilities should probably be noted.

### **General comments**

In February 2008 the SRC recommended that the monitoring team use seasonal moving averages, as the monitoring team had presented them in the February meeting. This report did not do that.

The SRC had recommended that caveats be provided on the limitations of the data. However, I did not see much discussion in this report on data limitations. I think that where a discussion of data limitations is most needed is on the use of estimates of days since death in estimating mortality. The implied uncertainty in the estimates of days since death was huge for most fatality records, so I'm surprised the monitoring team made no effort to discuss their use of these estimates.

The SRC had recommended that the monitoring team compare the number of birds seen during point counts, or relative abundance, to fatality data. However, the monitoring team compared minutes of activity, though I'm not sure this is what the monitoring team really did. If it was minutes of activity, that is not relative abundance or numbers of individuals, and it was not the type of unit one normally derives from point counts.

The SRC asked the monitoring team to estimate mortality separately for small, medium, and large turbines, because these classes of turbines were not all randomly selected by WEST, Inc. However, the monitoring team appears to have treated all turbines the same. Furthermore, there was no mention of how data from East Bay Regional Parks District were used. In my opinion, having analyzed those data, I don't think the monitoring team should have included those data in its mortality estimates. Also, there was no mention of Buena Vista or Northwind energy.

The SRC and monitoring team had agreed in February that an SRC subcommittee would work with the monitoring team to complete the report. The sub-committee was not asked by the monitoring team to help with this draft, other than many questions directed my way about data base issues. The sub-committee could have checked that the monitoring team was addressing the SRC's recommendations. I offered comments when I did hear about methodological changes, but I felt like my comments were not of interest to more than one member of the monitoring team. I feel that the SRC should have been informed of the monitoring team's intention to use estimates of days since death, because I feel that the experienced biologists on the SRC could have provided some useful feedback on this method.

Finally, I am concerned about the posting of the data used by the monitoring team. When I worked with the monitoring team on the data base in July, I found some significant problems with the data. I feel good about noticing these and helping the monitoring team to clean up the data set, but I have to admit that I doubt all the problems were fixed. The discrepancies in numbers of fatalities and numbers of MW compared also worry me about the posting of the data. I think the data should be examined more carefully before being posted.

**Joanna Burger**

## COMMENTS ON THE MONITORING REPORT

OVERVIEW: I appreciate the time and consideration that went into this report. The subject is complicated by a number of issues, not the least of which is the changing mitigation measures, the difficulty of defining mortality rates in the absence of population data, and the problems of comparing present data to a baseline.

My overall comments revolve around 4 issues:

1. The importance of clarifying at each point the data being presented, and if it is a rate, what the rate represents.
2. The importance of having sufficient discussion of the caveats to the data, methodology, analysis, and conclusions.
3. The importance of resolving the differences in final mortality estimates for the data as the Monitoring Team computed it and Shawn's methods.
4. The description, and indeed the execution, of the point count data is confusing, difficult to follow, and it should be clear exactly what was done, and why. There is a vast difference between ten birds and one bird seen for each of ten minutes.

## MY SPECIFIC COMMENTS AND QUESTIONS

1. 1.1: I am concerned that the report is already a year old, and some people will miss that these data only went through September of last year (2007). This should be clarified.
2. 1.2 Rate should not be used, especially in the executive summary unless it is defined.
3. 1.2: I cannot agree with the statement as written "that the mortality rates of the four target species increased" You need to make clear if you mean each species separately, or all four together.
4. 2-3: I would have the beginning sentence refer to the actual mortality that is the target (e.g. 50%).
5. 2-3: Again, in the second paragraph, the dates of the current study, and of the baseline should be stated.
6. 3.-4 and table. There should be a little more explanation of the boxes.
7. 3.-7. Each method should be justified, including why different methods were used one time compared to another.
8. This method section: it should be clear that the study did not examine relative abundance, but use as measured by presence in each minute.
9. 4-11: Can there be some measure of variance on the graphs?
10. 4-12: again, can there be some measure of variance on avian use?
11. The above comment goes for many of the graphics. How did the variance change by study plot and by time within each month?
12. Discussion: This section needs some work in that there are many issues that need to be at least mentioned, and some of them discussed. It is rather surprising to have such an increase in mortality, given the mitigation measures, and some discussion of this in terms of both reality, measurement error, design error, and other considerations needs to be stated. At the least, there should be some discussion of the limited time of the study (it does not include the last year).

13. I would feel more comfortable with there being some discussion about the species differences, confidence of the team in the data, and what they mean in terms of monitoring and future work

**Jim Estep**

**Altamont Pass Wind Resource Area Bird Fatality Study, July 2008  
Comments – Estep (August 20, 2008)**

Initial General Comment

Overall, the report is well-written with a very readable report structure and design. It is a significant improvement over the previous version.

Cover and Title Pages

Recommend changing the name of the report as follows to reflect its purpose and sequence in the monitoring process:

Altamont Pass Wind Resource Area Bird Fatality Study  
Annual Report, Years 1 and 2  
July 2008

Executive Summary

The first paragraph of the ES should also more clearly state the nature of this report and its sequence in the process by indicating that it is the annual progress report that summarizes data from Years 1 and 2 of the study.

I think it might also be a good idea to state up-front that the methods and results included in the document remain issues of discussion and concern by the SRC and that future progress reports will continue to reflect the results of these discussions and incorporate SRC recommendations.

Forth bullet: the term ‘local abundance (bird use)’ is a bit confusing. Abundance and use are two different things and this suggests that the terms are being used here interchangeably. Also, the bullet indicates that behaviors were analyzed. The results in the report do not present any data or discussion of bird behavior and the relationship to turbine-related mortality.

The conclusion statement might also be modified to reflect the interim nature of these conclusions and the ongoing debate regarding methods and analysis. The conclusion statement (i.e., “In conclusion, the study indicates...”) may be overly definitive. Might modify with “the results to date suggest” or something along those lines.

I would also recommend including an additional paragraph at the end of the ES as well as in the Discussion that describes the limitations and/or ongoing discussion regarding study methods and data analysis.

Section 2. Introduction and Study Area

Page 2-1. First paragraph, last sentence.

I would briefly describe applicable laws and regulations (i.e., CESA, ESA, BGEPA, MBTA, Fish and Game Code) in order to provide a little context to the uninformed reader.

Page 3-3. 3.1.2 Baseline Study

Might expand this description to include the differences between the number of searches conducted between the current study and the CEC study. This may be a limitation of the comparative analysis that should be acknowledged.

Page 3-4. 3.2 Bird Use Monitoring

There is some confusion here about how ‘abundance’ is defined and its relationship with bird ‘use’. The method that was used (1-minute intervals for 10 minutes) is a little different than a typical point count survey, which uses raw count data collected continuously over the survey period to estimate relative abundance.

There are variations, including some that are interval-based, used to establish time-of-detection and can then be used in a sort of capture-recapture model. But the terminology should be clarified and more detail provided for the data collection and analysis methods, including citations for the methods used.

As it currently reads, it appears that the way the 1-minute interval technique was used allows for multiple counting of the same bird, which might be useful in terms of evaluating use, but can create problems for relative abundance estimates. Generally, when attempting to calculate relative abundance through point count surveys, attempts are made to avoid multiple counting of the same bird.

Page 3-12. 3.4.6 Bird Mortality and Bird Use

So it would seem that the estimates derived from bird observations were based on bird use, not relative abundance. This might be interesting and useful, but the SRC was interested in investigating the relationship between changes in relative abundance over time and mortality.

Page 4-4. First paragraph

Given that approximately twice as many records were excluded from the baseline study vs. the current study, it seems like this should warrant additional explanation in terms of the specific reasons and any consequences in terms of potential limitations in interpreting the results.

Page 4-4. Table 4

Why are golden eagle and red-tailed hawk rows repeated in the table and referred to as ‘continued’?

## Page 4-11. 4.5 Bird Use

Same issue here as above. It's a little unclear what is actually being correlated. It's not relative abundance, but instead an estimate of bird use. I'm not sure what this is really telling us (the longer a red-tailed hawk remains in a given area, the greater likelihood for collision mortality?). I think the SRC was more interested in investigating correlations between the changes in seasonal abundance of different species in the APWRA and mortality.

## Figure 9

I'm a little confused by Figure 9. The text above it describes it as "...average monthly number of target raptor species observed". January '06 shows about 270 red-tailed hawks recorded during point count surveys. Or at least I'm assuming this is the total number recorded for that month (and not the mean number of observations per 10 minute interval as the table suggests). If so, what's being averaged? How are these numbers derived using mean obs/10 min.? I'm sure it is explained easily, but I'm having difficulty seeing it.

## Figures 10a – 13a

Why don't the monthly totals for the four target species in Figures 10a, 11a, 12a, and 13a match the monthly totals in Figure 9?

Also, while these figures are fairly self-explanatory, I would like to see some additional text explaining the relationships in each figure. Figure 12b in particular could use some explanation given it is a single outlier point that determines the relationship in this case.

## Page 5-1. Discussion

Given that this is an interim annual report and the limitations in the analysis, I would limit the definitive statements regarding the mortality estimates (e.g., "These results clearly show that...").

The discussion section is very brief. Given the interim nature of this report, I don't have a problem that it is brief; however, it would seem that there are relevant issues to discuss at this point, particularly since the results suggest a fairly dramatic increase in mortality. I would also include a discussion that describes the limitations of the analysis and describes the changes that will be implemented (e.g., QA/QC study) that might help to improve the analysis.

Review of the July 2008 Monitoring Report (M21)  
 By Julie Yee  
 August 18, 2008

Here are *Report Comments* based on my review of the Monitoring Report (M21) released July 2008. After the SRC discussed this report with the Monitoring Team during the conference call on August 14, 2008, I had further thoughts about some of the questions that were raised during the call. These include the use of midpoint for determining days dead, the reason for conducting detection trials (or not conducting as the case was in the early part of study), and the modeling approach for statistical inference. So, in addition to *Report Comments*, I have expanded on the *Midpoint*, *Detection*, and *Modeling* issues.

### ***Report Comments***

Page 1-2. Bottom of 1<sup>st</sup> paragraph. “This bias likely resulted in an underestimation of burrowing owl use and an inflation of the use/mortality correlation value.” Avoid making this statement unless you also discuss this in the report. Also change “correlation” to “rate” if that is what was intended. Otherwise it makes no sense.

Page 1-5. The comparison with Diablo Wind turbines in Table ES 3 is interesting, but I’d like to see the potential confounders addressed or at least mentioned. For example, the two sets represent separate parts of the Altamont. Also, would bird strikes at the larger turbines tend to fall farther away from the turbine? A bias will occur if turbine-related fatalities are falling outside the search radius at a higher rate at Diablo turbines compared to other turbines.

Page 3-4. Bottom of 1<sup>st</sup> paragraph. Bird use monitoring was conducted during the Baseline Study, so the statement is misleading.

Page 3-5 (Figure 4) What happened to incidental finds (WRRS fatalities)?

Page 3-6 (Figure 5) I have trouble following this flow chart. Do the item numbers (i.e. Turbine Strikes 1a, 1b, 1c, 1d) correspond to anything meaningful for the reader? What is meant by “conflicting information”? What is meant by “Other turbine related cause” [decision flows to this box only when “Turbine related injuries” was No in the upper right box – this seems contradictory]?

Page 3-8 (Figure 6) Why are there two different back date ranges for certain carcass conditions (i.e. 0-3 or 4-7 days for fresh flesh, and 8-30 and 31-60 days for gooey flesh)?

I appreciate that there is some attempt to backdate the age of the fatality, and I trust the experienced biologists to have the knowledge to assign a reasonable range of days dead as well as to admit when the range cannot be reliably made more precise. I can accept a biological judgment that a carcass is 0-90 days dead. Where I get worried is when a precise date for days dead is assumed (i.e. such as a midpoint of 45 days) in place of the full range. Also, the midpoint assumption was never even stated in the report. The report needs to further explain the decision rules leading to the different ranges of backdates. Also, what midpoint was defined for

0-3 days, 4-7 days, and 0-7 days? As for the midpoint assumption, see *Midpoint* comment below.

Page 3-9. 2<sup>nd</sup> paragraph. There were no means or standard deviations for unadjusted fatalities are reported. Remove description if no longer current.

Page 3-9. equation. Put parentheses around “RxP” otherwise this expression “M/RxP” would conventionally be interpreted as (M/R)xP which is not correct.

Page 3-9. top of last paragraph. As a clarification, the SRC did not request to not conduct searcher detection trials. Some information on searcher detection rates was already available, and it was not a priority to conduct further study until this year. See *Detection* comment below.

Page 4-3. There is a striking difference in the proportion of records retained between Baseline (64%) and Current study (83%). Someone should examine the data to assess whether the excluded records followed any patterns that could contribute to biases. Perhaps the exclusion rules need to be revisited. Also, I would rather see these numbers tallied for the four focal species or separately by focal species, rather than all birds combined. (I would even suggest doing this before changing any of the exclusion rules).

Page 4-6. I’m not completely comfortable with this ANOVA. The nice thing about it is that it compares the two study periods in a straightforward manner. But this is an unweighted ANOVA, so all months are weighted equally despite the large variation in monthly effort. According to Figure 8, the effort ranged 2.4 – 106 MW/month during the Baseline, and 77 – 282 MW/month during the Current study. The low effort months are subject to higher variance and more outliers because of the decreased likelihood of finding a fatality; and when a fatality is found, the per MW rate spikes. The high effort months should be more stable and therefore arguably should have greater weight in determining the mean. However, if the monthly rates are weighted, then this could lead to disproportionate weights among seasons or years. If this is the case, then this could be a concern because of the potential bias associated with placing uneven weight on high vs low mortality seasons, or high vs low mortality years. If there’s going to be an ANOVA, then I at least request that the residuals be examined for gross violations of the ANOVA assumptions. The report need not contain a full residual analysis, but at least document whether the ANOVA assumptions appear satisfactory. I would not discourage using the ANOVA (unless severe biases were discovered). But I would have more confidence if the data were analyzed by other models based on different assumptions, and the results showed robustness to the analysis approach. See *Modeling* comment below.

Pages 4-8 through 4-10. It would be valuable to have standard errors and confidence intervals on those final statistics on percent change (Tables 5-7).

**Midpoint Issue**

There was a lot of concern expressed during our Aug 14 conference call about the differences between Shawn Smallwood’s estimates and the Monitoring Team’s July 2008 estimates. This issue is complicated, but I contemplated all the differences (which boil down to three main features, below) and found I prefer a combination of aspects from both estimation approaches. By assessing a separate  $R_C$  correction for each fatality, the monitoring team appears to have followed part of my recommendations from my review of the February 2008 report (document P80, comment #4). I had recommended using the search interval that corresponded to the fatality on a case by case basis. For example, if a fatality was found following a 40-day search interval for that particular location, then adjust the fatality using  $R_C$  based on 40 days. I was not fond of applying a blanket correction based on 37-day or 44-day average search intervals, and I still am not.

There is another difference, in which the Monitoring Team applied their correction to individual unadjusted fatalities prior to summing the resulting adjusted fatalities, in contrast to Smallwood’s approach in which he summed the unadjusted fatalities before applying the correction to get the adjusted total. When the intervals are equal, then the outcome of both calculations are identical [i.e.  $\sum_i (X_i/R_c) = (\sum_i X_i)/R_c$  ], so this is a relatively minor difference compared to the issue of what to use for  $R_C$ .

Summary of differences between two approaches:

Smallwood approach, and MT Feb 2008 draft report	Monitoring Team July 2008 report
<ul style="list-style-type: none"> <li>• Add fatalities, then adjust rate</li> <li>• Use average search interval</li> <li>• Intervals are either 37 or 44 days (see MT Feb 2008 report)</li> </ul>	<ul style="list-style-type: none"> <li>• Adjust first, then add adjusted rates</li> <li>• Use midpoint of range of days dead</li> <li>• Typical intervals include: 45 (when 0-90 or 31-60 days dead) 60 (when 31-90 days dead) 19 (when 8-30 days dead)</li> </ul>

The MT used the same function for  $R_C$  as found in Smallwood (2007), where the only difference is that MT used “Days Dead” as their look-up value and Smallwood used “Time since last search ( $d$ ).” The function was derived by Smallwood and works under the assumption that the fatality occurred between 0 to  $d$  days prior to its discovery, with equal likelihood it occurred on any day between 0 and  $d$ . For example, for a 37-day search interval, then implicit in Smallwood’s method is that the Days Dead falls within the range of 0-37 days. So the notion of Days Dead is present in both approaches, although Smallwood does not call it that. The MT is doing nothing new by assuming a range of Days Dead. But the MT altered the procedure in two very significant ways. First, they have allowed some very short ranges and some very long ranges (i.e. 0-3 days and 31-90 days) to incorporate additional information on carcass condition (or lack thereof) and the possibility of searcher error; this could be a great enhancement to the procedure after developing consensus among biological experts on the range classification. The second change is: they took the midpoint of the range, presumably as an unbiased guess for the

number of days the carcass was in the field. This is a problem because  $R_C$  works as a function of upper range, not a function of estimated actual days in field. So, for example, when a carcass was estimated to be 0-90 Days Dead and the MT set Days Dead = 45, then they got instead the correction for a carcass that is for 0-45 Days Dead. As another example, when the carcass is estimated to be 31-90 Days Dead, then the midpoint is 60 days and the correction is actually for 0-60 Days Dead.

I agree with different aspects of both Smallwood and MT approaches. I agree (with MT) with adjusting fatalities on a case-by-case basis, because different fatalities are associated with different ranges of plausible days since death. I disagree (with Smallwood) on the use of average search interval, but can accept using search interval in some way to determine the correction  $R_C$ . On the other hand, I realize that some fatalities can be older than the interval when the area was last searched, and can accept the classification of fatalities into intervals of up to 60 or 90 days. I disagree (with MT) on setting days dead to the midpoint of large intervals, and I believe this was a misuse of  $R_C$ . Smallwood's derivation of  $R_C$  makes it unnecessary to estimate Days Dead to the day.

Finally, referring to a comment made in the Aug 14 conference call, I agree (with MT) that further information about searcher detection and carcass removal processes could help inform a more appropriate way to determine an  $R_C$  type of adjustment.

### ***Detection Issue***

In the July in-person meeting, an audience member asked why there was 0% confidence in the current approach on correction factors (including detection), and I gave what I felt was an unsatisfying response by saying that the FPLE statistician Bill Warren-Hicks would not accept the current approach. Meanwhile, the MT stated in their report that the SRC had requested detection surveys not be done. I believe that the public and MT do not understand how the decision for the SRC to recommend detection studies at this late stage evolved, and I would like to set the record straight (speaking for myself).

At the time the SRC began meeting, the monitoring protocol was underway. One year of data had already been collected, and the only aspect of the survey that I felt the SRC had the influence to modify was the number of turbines. We discussed the issue of detection and scavenger errors at length, and whether or not it was even necessary to adjust raw fatality rates (i.e. why not compare average raw counts between the Baseline and Current studies?). There were two reasons I found most compelling for continuing to develop the adjustment rates. Firstly, the Settlement Agreement was focused on 50% reduction of estimated total mortality of the four focal species, so we could not simply look at raw fatality counts. Secondly, the different lengths of search intervals between the Baseline and Current study, combined with scavengers' effect on reducing raw fatalities, had strong potential to bias the comparison. We discussed detection error, but I viewed this as less of a source of bias than scavenger error, because there was no obvious reason to expect the detection error to be any less appropriate for the Current Study than it had been for the Baseline Study. If the detection errors were biased because they were produced from outside surveys, then they would be equally biased for both the Baseline and Current Study. If our goal was to compare apples to apples and look for 50% reduction in size, then there was no compelling reason to consider proposing an additional detection study. When Bill Warren-Hicks criticized the current approach, he made two main arguments. The first argument was that the detection rate is dependent on the scavenger removal rate, and the two

rates cannot be estimated independently as they were in the current approach. I continue to disagree with this argument. We exchanged discussions on this for some time and could not get one another to budge from our respective views [While I agree that marginal probability of detection is dependent on the probability of scavenging, I also maintain that the estimator in question used for the correction factor is the conditional probability of detection, given the carcass is unscavenged, and that this estimator is not dependent on the probability of scavenging]. His second argument was that the scavenger removal and detection surveys were from outside studies and could not be reliably applied to the Altamont. Nobody disagreed, and there was no better way to resolve that issue than by developing the QA/QC study. I continue to have reasonable confidence in the current approach, but it relies on assumptions that FPLE has challenged and which cannot be resolved without conducting the independent QA/QC survey in the Altamont. I believe the current approach is acceptable for purposes of comparing Baseline to Current Study, but I also believe that the Current Study with the QA/QC approach is a more ideal survey because it will enable a more reliable estimate of current rates of fatalities. This would be critical if the Current Study were to ever become the new baseline.

### ***Modeling Issue***

The analyses presented by the monitoring team so far are sample-based. In other words, they are statistically valid only because the turbine sites were randomly selected and evenly sampled each month across exactly two years. When the sampling is performed unevenly, then this introduces the possibility that certain parts of the year or certain parts of the Altamont are more heavily represented than others. The Current Study is less vulnerable to this than the Baseline Study, which had much more uneven sampling across space and time. This concerns me somewhat. I would prefer that the comparison be explored via model-based inference as well as sample-based. The advantage of a model-based inference is that it can be made more robust to differences in how the data were sampled, and can be used as a check on the sample-based analysis. I don't have a model proposal for the MT, but I did roughly outline some ideas in my last review (document P80, item #8).

**Sue Orloff**

Comments on the July 2008 Monitoring Report (M21)

By Sue Orloff, August 25, 2008

I have read over all the comments by other SRC members. Since many of my concerns have already been addressed by them, I will only cover those issues that have not.

Page 3-7 (3.3.1 second paragraph): The first two sentences appear to contradict each other. The first sentence says that the Baseline Study removed fatalities that preceded the start date. The second sentence says Smallwood and Thelander studies (which I believe is what you are calling the Baseline Study) allowed fatalities to be backdated before the start of surveys. I think I know what you are trying to say but it is very confusing. Please clarify in text.

Page 4-6 (4.4.1): Julie has already said she has problems with the ANOVA and I agree with her concerns. In addition, I would like to see how each separate year of the current study compares to baseline. So far the current study shows a big difference between 2006 and 2007 in abundance (use?) and mortality. But you should probably wait until the last three months of 2007 are added to the data set to complete the year.

Page 5-1 (5.0 first paragraph): In the sentence “*These results clearly shows that, to date, avian mortality has not been reduced ...*” – besides not using words like clearly shows as other SRC members have suggested, it might help to use the word decrease rather than reduce. Using the word reduce in this sentence suggests that the mitigation measures have not had any effect on mortality. But we really don’t know that yet. Mortality could have been reduced by the mitigation measures but an increase in avian abundance since the baseline study could have resulted in increased mortality. We need to factor in abundance data first.

Lastly, in the new section on limitations of the data (which everyone has suggested you add to the report), please include a discussion on the large annual variability in abundance and mortality (as shown in the current study) and how this will affect the comparisons to baseline.