I&M Protocol Peer-Review

Protocol Title: Whooping Crane Winter Abundance Survey Protocol **Version:** 1.0
Survey Identification Number: FF02RTAR00-002
Refuge: Aransas National Wildlife Refuge
Authors: Matthew J. Butler, Biometrician, U.S. Fish and Wildlife Service
Bradley N. Strobel, Wildlife Biologist, U.S. Fish and Wildlife Service
Cinthia Eichhorn, Regional Data Manager, U.S. Fish and Wildlife Service

Protocol Review Timeline:

6/15/13 – Protocol submitted for peer review.

9/23/13 – Peer reviews received and returned to authors.

Note: Authors notified of unanimity, and reviews not in order (3 external to USFWS (Joshua Schmidt, Charles Anderson, John French); 2 internal to USFWS (Beau Hardegree, Wade Harrell)

3/26/14 – Protocol revision received from authors.

4/22/2014 – Final decision (accepted).

Letter From Dr. Harris:

22 April 2014

Dear Dr. Butler:

I have carefully reviewed the "Whooping Crane Winter Abundance Survey Protocol" and the response to reviewers. Clearly, the reviewer comments were addressed thoroughly. By doing so, this protocol revision has improved in content and clarity.

The work done by you and the co-authors is much appreciated by the United States Fish and Wildlife Service, and will advance the understanding and conservation of this whooping crane population.

Please consider this protocol accepted and finalized.

Regards,

Grant Harris

Letter to Dr. Harris:

26 March 2014

Dear Dr. Harris,

We have completed revision of the Inventory and Monitory protocol entitled "Whooping Crane Winter Abundance Survey Protocol". Below we have attached the reviews and our response to questions, edits, or suggestions raised by each of the Reviewers. All comments from the reviewer have been addressed in this document (see blue text) or within the protocol.

We have included some R code in appropriate places to help demonstrate analyses (see Reviewer 1, comment 2) or support simulation efforts to better demonstrate the behavior of Generalized HDS models (see Reviewer 2, comment 14). The R code throughout this rebuttal is in `Lucida Console` font.

We believe this critique has helped us craft a better protocol. We appreciate and thank each Reviewer for their time and effort.

Sincerely,

Matthew J. Butler

Letter From Dr. Harris:

23 September 2013

Dear Dr. Butler:

I have received 3 external reviews and 2 internal reviews of your protocol titled "Wintering Whooping Crane Abundance Survey". All Reviewers identified points worth addressing in a revision. Notably, I think more effort could be placed on the following:

1) Further justification and description regarding implementation of the distance sampling technique.

   We have provided a new section in Element 1 entitled "Justification of objectives." We have also provided additional discussion of the distance sampling technique throughout the protocol in appropriate places.

2) Reducing some discussion of the former method, keeping only the salient deficiencies; namely those necessary to describe why a revision and new approach was required.

   Some description and critique of the traditional survey technique is all we provided (i.e., less than 4 pages of a document containing over 150 pages). As we began addressing the Reviewers' comments, we have begun to wonder if we spent enough time critiquing the traditional technique since many of the Reviewers' comments have highlighted their misunderstanding of the traditional technique which further emphasizes the need for critiquing it (we point out each of those comments below). Many of the Reviewers revealed several misconceptions about the traditional technique. Therefore, we have chosen to include the description and critique of the traditional survey technique. In a few cases we had to expand the critique to address some of the Reviewers' questions (e.g., misconceptions about the enumeration of mortalities during the traditional surveys).

3) Please increase size of text in Figure 1.

   We have made all map-based figures larger (e.g., Figs. 1, 5, 6, and 8). Originally these figures were only 4.5 inches wide but now they are 6.5 inches wide. We are unable to increase the font size associated with the background layers within these figures because those fonts are embedded as part of a basemap in ArcGIS. However, by increasing the size of the figures, the text is much larger. After increasing the size of the map-based figures, we decided to increase the size of all figures.

Reviewers provided general and specific comments. Please respond to each line by line, and identify where you modified the protocol as/if necessary.

Below we address all comments from the Reviewers in blue and note the section in the protocol in which changes (if necessary) were made.

Once this document and revised protocol are returned, I will evaluate your responses and revision. At that point, the protocol may be declined, accepted or further reviews sought.

Sincerely,

Grant Harris
Chief Biological Sciences
United States Fish and Wildlife Service, Region 2
Albuquerque, NM 87103

REVIEWER #1:

Overall, this is a much-needed, thorough and excellent upgrade to the survey methods for whoopers at Aransas. I am impressed with the amount of effort and thoughtfulness evident in this review draft.

Thank you for your thoughtful review; we value your comments and appreciate your efforts. We want to note upfront that this survey extends well beyond the borders of Aransas National Wildlife Refuge and covers over 1,200 km^2 of the winter grounds. There seems to be a misunderstanding that whooping crane ostensibly overwinter only on Aransas National Wildlife Refuge. However, 55.3% of new primary sampling frame covers areas beyond the boundaries of Aransas National Wildlife Refuge. We have included the boundaries of Aransas NWR in Figure 6 to help illustrate this point.

General comments:

1. Too much time spent critiquing the former survey method; this is evident in the description multiple times of the new methods as "defensible", "scientific", etc.

   We felt it important to make clear that the new technique is "defensible" and "scientific." The previous methods were not, hence the need for this protocol. We have removed some of this language as it was repetitive.

   Some description of the deficiencies is needed, as a platform for describing this method which will improve upon it, but the important point is the demonstrating the features, validity and benefits of the proposed method.

   Some description and critique of the traditional survey technique is all we provided (i.e., less than 4 pages of a document containing over 150 pages). As we began addressing the Reviewers' comments, we have begun to wonder if we spent enough time critiquing the traditional technique since many of this Reviewers' comments have supported the need for critiquing the traditional survey technique (there seems to be much misunderstanding about it). We point out each of those comments below. Other Reviewers had similar misconceptions about the traditional survey technique. Therefore, we have chosen to include the description and critique of the traditional survey technique as was originally included in the protocol. In addition, we have had to include some more information to dispel misconceptions about the enumeration of mortalities during the traditional surveys.

2. Related to (1), is the need to use the numbers derived from the former methods as we go forward with the conservation of whooping cranes. So those numbers cannot be described as useless, and

indeed they are useful (just not useful enough). A section is needed somewhere on how the data from the new methods can be compared properly to the numbers from the former methods.

We have never described the estimates derived from the traditional survey technique as useless and in fact we have used them to help better understand the dynamics of this population (see Butler et al. 2013). We undertook revision of the survey because the traditional technique needed improvements because 1) it was not repeatable and lacked standardization, 2) it provided no description of the uncertainty of abundance estimates, and 3) it was contingent upon an assumed ability to uniquely identify unmarked individual whooping cranes. Since the traditional estimates were not provided with SEs nor can those SEs be calculated, all one can do to compare new estimates with the old estimates is to continue to assume the old technique produced a complete census. We understand this is unsatisfactory, but there was no way to replicate the traditional technique with new observers to create some sort of correction factor.

Since the abundance estimates from the traditional technique have always be treated as if they are exact counts and unfortunately will likely continue to be treated that way, comparison between estimates from the two techniques is conceptually simple. For example, in 2010, the traditional estimate was 283 whooping cranes. In 2011, the new technique resulted in an estimate of 254 whooping cranes (0.126 CV). A simple *z*-test can be used to test if the 2011 estimate is different from the 2010 estimate ($z = -0.906$, $p = 0.182$; `pnorm((254-283)/(254*0.126))`). However, since we do not know how whooping cranes outside of the sampling frame were incorporated into the traditional survey technique, comparison with the new technique may be problematic since the new technique only applies to the primary sampling frame.

Finally, if we continue to assume the traditional technique produced exact abundances, then for trend analysis, weighted regression would be appropriate (Gerrodette 1991). The weights used would be the inverse of variance estimate for each abundance estimate; for the estimates from the traditional technique, 1 would be used as the weight (assuming no uncertainty in traditional estimates). Of course this makes many uncomfortable, but how else can uncertainty about the past estimate be incorporated when uncertainty was not quantified. Perhaps, assuming no uncertainty in the traditional estimates is a poor assumption but we have no other estimate of the uncertainty. We could be more conservative, and give them less weight by assuming a CV of 15% or some other value, but that would be an arbitrary decision.

3. Figures are excellent; very helpful (although Fig 1 would be better larger…).

   Thank you, we strove to provide useful figures that described the issues. We have made all map-based figures larger (e.g., Figs. 1, 5, 6, and 8). Originally these figures were only 4.5 inches wide but now they are 6.5 inches wide. We are unable to increase the font size associated with the background layers within these figures because those fonts are embedded as part of a basemap in ArcGIS. After increasing the size of the map-based figures, we decided to increase the size of all the figures in the protocol.

4. A couple of references were made to later revisions of the protocol; did I miss a calendar for revision?

   We considered including a calendar for protocol revision but decided protocol revision should occur on an "as needed basis" as recommended in the Survey Protocol Handbook (U.S. Fish and Wildlife

Service 2013). We added some additional description of the revision process to the "Revision History" section.

Related to that there are likely to be practical limitations on carrying out some of the specifics; those kind of could/should be changed before the date of a more complete revision.

We are unsure what "practical limitation" is being referred to. The protocol, as it is currently designed, can be carried out as planned. It would be poor design on our part if the protocol could not be implemented in the field.

5. Definition of transect; it seems to be used as both the flight path of the aircraft and the width of territory sampled by the flights: choose one, define it and stick to it (I suggest the width of territory sampled…).

   Just as with any other line-transect based survey, the transects are the lines from which the survey is conducted. The transects are fixed-width strip transects. Therefore, the transect refers both to the line and the 500-m strip on either side of the line. It is very much appropriate to use the word transect for the flight path or the strip around the flight path. We defined the transect in "Element 3, Sampling units" section and "Element 3, Transect spacing". The transect is the flight-path and the survey area is a fixed-width around each transect. The transect configuration is independent of whooping crane territories.

Specific comments:

1. P2, L19 – I was under the impression that the flight paths were a more consistent distance apart…

   Though erroneously so, apparently many were under the impression that flight paths were closer together and more consistent during the traditional survey. Perhaps this was because Stehn and Taylor (2008) reported that transects were spaced 400 to 500 m apart. However, GPS track data from the 2010 surveys indicated transect spacing varied between 250 to 800 m. Unfortunately, no GPS track data were saved by observers during the previous years. The traditional surveys were much less systematic than portrayed in Stehn and Taylor (2008) and were more targeted to areas known to be occupied by whooping cranes. We believe this is a good example of why we needed to critique the traditional survey technique.

2. P2, L24 – define briefly 'survey blocks' in this sentence.

   We have changed the sentence to read: "Survey duration was typically between 5–6 hours with a rest break approximately half-way through the survey during which the flight crew returned to the Aransas County Airport." We see no reason to define the concept of a survey block at this point since the traditional technique did not have survey blocks delineated and defined as the new technique does.

3. P4, L35 – consider deleting this; I suspect that Tom could identify many of the cranes.

   The statement in question is "The observer assumed ability to uniquely identify unmarked individual whooping cranes." It is true the observer assumed he had this ability though he provided no evidence that he could. We consider an assumed ability to identify unmarked birds without any

uniquely identifying characteristics untenable. In our professional experience, defensible identification of free-ranging individuals requires the use of unique markers. While some believe that Tom Stehn could consistently identify unmarked whooping cranes there are no data available to support this belief. As such, we feel our comment is accurate. This forms another good example of why we needed to critique the traditional survey technique.

4. P4, L39 – is there a citation for this detection function, in another birds or situation?

   On this line we stated that "group size was likely positively related to the probability of detection." Size-biased detection is very common in these types of surveys and has been noted in many species, both birds and mammals. We have provided some citations where detection probability was related to group size.

5. P4, L42 and L43 – are you sure this is true, at the transect width that was used formerly? I was under the impression that the transect width was narrow enough that detection did not vary across it… maybe not.

   Yes, we are sure this is true. See Strobel and Butler (2014) for detailed analysis of detection probabilities from the traditional survey technique. They show that detection probability of a whooping crane group within 500 m of the survey line was 0.558 (SE = 0.031). We summarized these results in "Element 2, Winter 2010–2011 surveys" section. Even if the traditional survey technique used 400 m wide transects (200-m half-width), detection probability was still less than 1.0. The detection function in Strobel and Butler (2014) shows that the detection probability within 200 m of the survey line was 0.69. Narrowing the transect does not ensure complete detection. Again, this emphasizes the need for critiquing the traditional survey technique.

6. P5, L21 – define 'surveyed area', or refer to the later section where it is defined, and precision is a VERY important aspect of this or any survey; a review of the justification of "...enough precision to detect a 10-15% decline over 3 years" is needed, and probably warrants a separate section; I am not sure that is adequate (although it may be all that is practical), but I could be convinced; I <u>am</u> sure that some on the WCRT will object to that criterion as too lax, and I am also sure that there is a decidedly false and overblown sense of precision in the numbers form the current methods. This issue needs to be met head on, and can be discussed usefully with reference to the costs outlined later in the document.

   We added a reference to the later section for the "survey area."

   We added a new section to summarize the justifications provided in Butler et al. (2013) for developing a survey with enough precision to detect a 10–15% annual population decline over 3- to 4-year period. See new section, "Justification of objectives" in Element 1.

   We agree that "there is a decidedly false and overblown sense of precision in the numbers form the" traditional technique. That is why we believe the critique of the traditional technique is so important. We do not think there is an overblown sense of precision with the new technique. Early on precision has been lower than we wanted (12% to 19% CV) but now that all the kinks have been worked out of the protocol, we expect better precision from future surveys (see discussion of Internal Reviewer #2's general comment #2).

7. P5, L24 – I realize you don't want to imply breeding pair, or any other biological interpretation to these associations, so create operational definitions of paired birds, and recruitive pairs (an awkward term), and leave out the quotes.

   An operational definition was already provided parenthetically after each term. We removed the quotes from these two terms.

8. P9, L6 – good section: what about the type of aircraft (airspeed, nose, comfort, etc.)

   We have added some additional information about the type of aircraft required for this protocol to "Element 4, Aircraft type" section. Aircraft speed is specified in "Element 4, Field data collection methods" section and was based on the traditional survey technique.

9. P11 – very nice table and figure.

   Thank you. We have made all the map-based figures larger.

10. P12, L8 – here might be a place to indicate how the data collected with this new method will or should be compared to past numbers…

    See response to Reviewer's general comment #2. We have added a note about comparison of estimates from the two techniques to "Element 3, Comparison with estimates from the traditional technique."

11. P12, L22 – do you mean PSF…?

    The sentence in question is "Eventually, additional areas will need to be added to the SSF as the population expands." We do not mean PSF, we mean SSF. The SSF regions will be promoted into the PSF when criteria outlined in "Element 3, Sampling frame" section are met. Then new regions will be added to the SSF once SSF regions are promoted to PSF regions. We reworded this sentence to clarify: "As SSF regions are promoted into the PSF as the population expands, additional areas will need to be added to the SSF."

12. P12, L21 and L22 – also should use anecdotal information to add areas for formal survey (secondary); one does not want the methods designed to survey birds where they are now, to constrain the places one looks for cranes as they expand: the birds may get to very different areas (this applies to the habitat modeling efforts too, probably more so).

    We speak to this in the "Element 3, Sampling frame" section. "Because the Aransas-Wood Buffalo population is growing at ≈4% per year, we will address how to add more SSF regions when this monitoring plan is due for revision. For example, the spatially-explicit models of habitat use could be used to predict areas composed of the best potential habitat." We did not suggest other anecdotal information could potentially be used as well so we have added this sentence. "Ancillary information gained from internal and external sources (e.g., fortuitous public or staff observations; Texas Whooper Watch) could potentially be used to identify new SSF regions as well."

13. P13, L18 – I suggest the numbering/lettering system go from the shore north: it is easier to add survey regions/blocks which will probably be inland.

Survey blocks are lettered with the survey blocks initials (i.e., Blackjack Peninsula is labeled with BJ). Since there is no plan to add additional transects to specific survey blocks but instead add new survey blocks as the population expands then we see no reason to change the way in which transect are numbered within a survey block. Besides, since all survey information is now spatially explicit unlike before, we can simply add new transects as needed and renumber them appropriately. Since all the transects are displayed on the computer during the survey, renumbering them if additional transects were added would not cause confusion.

14. P13, L30 – the term peak abundance is a very poor one, and should be discarded in favor one without interpretation imbedded in it; I am very interested in what the greatest number of birds at Aransas may be, but there may be/are many opinions on how to get that number; it is interesting that numbers from past survey methods are the highest in Dec, but that might not be the case with the new method; better might be to define the period intense survey with reference to the dates of migration (it might be some function of departure date, arrival dates, etc.); this would allow the dates of intensive surveys to vary with the birds; I can imagine that seasonal differences may alter the date of 'peak abundance' and indeed climate change may well alter the phenology of crane migrations over the years; call it the "early resident period"??

    We agree that the peak abundance term may be imbedded with too much interpretation and have removed it from the protocol. There is no reason to expect this method to result in higher abundance later in the winter. To ensure comparability from one year to the next, surveys need to be conducted during the same time period. To facilitate comparison with the traditional technique, which aimed to provide estimates of "peak" abundance (i.e., abundance post fall-migration prior to any winter mortality), then December is the most appropriate time period.

    We know of no way to rigorously tie survey timing to anecdotal observations of migration phenology. Based on the historic data, we felt that surveys probably could begin either the second or third week of November. However, we delayed starting surveys until December just in case migration was delayed some years. We agree that if we see that migration is delayed in a particular year, then we as program managers need to delay surveys a little while for that year. In fact, that is part of the reason we allowed surveys to be conducted anytime during a 1-month period (as long as they are blocked together into a 2-week window). If a trend of later migration dates becomes apparent, then of course we would need to re-evaluate the survey dates for future surveys.

15. P14, L28 – I can well imagine (but do not know firsthand) that habitat use will vary thought the season, and especially for the birds of the year; hence spatially explicit models relevant to Feb-Mar might be very revealing.

    We agree that the spatially-explicit models of abundance might provide interesting insights into whooping crane habitat use during different periods of the winter. That is why we have provided a mechanism for developing such models during the secondary survey periods if the Refuge or stakeholders chooses to pursue surveys during January, February, or March. However, as outlined in "Element 3, Survey timing," surveys conducted during the secondary periods are optional and will only occur if clear objectives are identified by the stakeholders. We do not consider a passing curiosity about changes in habitat use through time a clear enough objective. Frankly, there are much better ways to understand temporal trends in habitat use such as satellite-telemetry.

16. P16, L3 – see general comment (4) above.

    See our response above concerning revisions of the protocol.

17. P17, L15 – very interesting…. And relevant to the discussion about precision, etc.; the numbers at Aransas can help us learn about life history during migration south, as well as help inform management needs at Aransas proper.

    Actually, the numbers at Aransas cannot inform us about life history during the southern migration because little information is available from the flock during the breeding season. Assuming the precision of estimates is tight enough, the work in Butler et al. (2013) shows that biennial monitoring would likely provide the information necessary to detect the kind of declines that would cause significant delays in time to recovery. But moreover, the discussion in Nicol and Chadès (2012) really reveals intensive monitoring when the policy decisions regarding downlisting are imminent is most important but other times are much less relevant.

18. P17, L31 – Unclear: is the TESC too coarse grained?

    It depends on the question one wants to ask of the data. If we wished to determine how whooping crane abundance responds to varying densities of blue crab across the landscape, then of course the TESC vegetation map is inappropriate. However, if all we wish to characterize is how whooping cranes distribute themselves among saltmarsh and upland habitats, then the TESC vegetation map is appropriate. So all we were trying to convey is that the spatial layer used in the analysis must be compatible with the question being asked. We find this idea is outlined clearly in the following paragraph from "Element 3, Environmental covariates" section of the protocol:

    "Hierarchical distance sampling results in spatially-explicit models of abundance that explicitly consider relationships between population density and environmental covariates while controlling for detectability (Royle et al. 2004, Chandler et al. 2011, Sillet et al. 2012). Currently, environmental covariates are derived from the Texas Ecological Systems Classification Project (http://www.tpwd.state.tx.us/gis/gallery/), Phase 3 (Ludeke et al. 2012; Appendix E). The percent of each 1-km2 transect segment is determined for 6 general vegetation types: saltmarsh, open water, wetland, saltmarsh-shrubland, upland, and urban (Appendix E). Though these covariates are only rudimentary, strong relationships with whooping crane abundance has been demonstrated with some of them (Strobel et al. 2012). As managers, biologists, and policy makers begin to recognize the value and potential of spatially-explicit models of resource use and develop more detailed hypotheses about whooping resource use, specific resource maps can be tailored to answer more specific management or policy questions. Maps from which environmental covariates are derived will need to be periodically updated with the most current data."

    We also note in "Element 4, Environmental covariates" section of the protocol:

    "As additional maps of vegetation layers, other whooping crane resources (i.e., salinity, blue crab abundance, etc.), or management actions (i.e., time since prescribed burn, fresh water provisioning, etc.) become available and more detailed hypotheses of whooping crane abundance relationships are developed, additional environmental covariates will be derived and incorporated into the hierarchical models of abundance."

19. P18, L19 – good; can an ID test be developed like is done for BBS observers? This would be useful. BTW, the emphasis on training and practice (later section) is excellent, but should be strengthened: you might consider requiring practice using the computer recording method, coordinating with the other observer (I like the idea of practice in a car..), and with identification; the BBS has shown very clearly how much difference experience, and hence training means to the quality of the data.

    We like the idea of an ID test but currently have no plans to implement such a test. We believe the training requirement that new observers must participate as a non-observer in the fourth seat of the aircraft during a minimum of 3 surveys is adequate training.

    We strengthened the language about practicing using the equipment from an automobile. Staff have implemented this practice session prior to beginning the surveys each year.

20. P19, L17 – is there detection variability by habitat of HY birds? (white birds are a bit easier..)

    There is likely some detection variability by habitat for all whooping cranes not just juveniles. However, as indicated in the section in question, groups with juvenile birds will have slightly higher detection probability than groups without juveniles simply because those groups with juveniles are larger groups.

21. P19, L36 – yes, good; and keep in mind that we have been consistently surprised by the habitat selection of the re-introduced whopping cranes, so as they expand in S. Texas, I now expect cranes in unexpected locations.

    Agreed, we already have seen unusual habitat selection as this population has grown (e.g., Granger Lake). By the way, we included a section in Appendix C entitled "More on prediction" that discusses prediction beyond the bounds of inference and possible uses of those predictions.

REVIWER #2:

General comments:

1. While I agree there is a potential for bias in the past estimates, the intensity and frequency of the surveys, combined with the small population size (with a large number marked during many years), suggests that the max estimates are probably pretty close to the true number of individuals, or at least a very good index. The relatively low annual variability in the estimates (Fig 2, page 3) suggests that this is the case. A poor index would result in a more 'jagged' graph due to sampling error.

   Unfortunately this Reviewer has little knowledge of the past intensity and frequency of the traditional surveys. No one but the previous observers even knew the intensity (i.e., what areas were surveyed) of their surveys since only information about detections were recorded (i.e., past observers did not record where they surveyed and did not detect birds). Further, the frequency of the traditional survey varied widely. The number of surveys conducted between December 1st and March 31st per year varied from 4 to 21 (mean = 12.7, SD = 4.2) indicating little consistency in survey effort from year to year (Butler et al. in prep).

Wild whooping cranes were not individually marked until 1975. In fact, only 132 marked birds have been in the population spanning the period of 1977 through 2004 (Gil-Weir et al. 2012). There were not "a large number of marked birds during many years." Sure, nearly 60% of birds were marked during winter 1988–1989 (Stehn 2004) but the proportion of marked cranes was typically much lower. Over 40% of the years in which the traditional survey was conducted (1950–2011), no color marked birds were in the population. By the winter of 2010–2011, only 7.8% of the flock was marked (Stehn 2011). It is also important to note that many of the banded cranes were not marked with uniquely identifiable color marks but with aluminum bands (Stehn 2011).

It seems that this Reviewer misunderstood the point of figure 2. Figure 2 shows how the traditional technique somehow resulted in an estimate that was often greater than the maximum number of birds observed but sometimes resulted in an estimate that was smaller than the maximum number of birds observed. This is disconcerting. If the observer observed 100 birds on a survey but only reported an estimate of 97 birds, we wonder what happened to those 3 birds. As we noted in the protocol, "this inconsistency may be attributed to the observer's interpretation or perception of double counts. However, we cannot know…since rules governing how data from separate surveys were combined or how public reports of whooping cranes from outside of the 'censused' area were incorporated was largely determined by the observer's opinion." We have developed this protocol to get away from a technique largely based on expert opinion and experience to something that can be replicated by staff that do not have many years of experience working on the species. As current staff move on to new jobs, we now have a technique that can be replicated by anyone.

Fundamental problems with the traditional technique can be illustrated easily. For example during the winter of 2008–2009, the traditional technique reported 270 birds on the wintering grounds with 23 of them dying during the winter. During the winter of 2009–2010, the population was reported to be 264 birds with 22 juveniles which would mean there were 242 adults in the population. However, the maximum count during the winter 2009–2010 was 270 birds which would presumably mean there were 248 adults in the population. If 23 mortalities occurred during the winter of 2008–2009, there could only be 247 adults in the population during the winter of 2009–2010 (270–23=247; assuming no mortality during other parts of the year). Where did that extra adult come from? Did the maximum count of 270 during the winter of 2009–2010 include 6 birds that were counted twice? Was mortality during the breeding- and migration-periods only 2% ((247–242)/247)? This is an illustration of the fundamental danger of a technique that portends to deal in absolute truth when there is really much uncertainty in the estimates.

2. I wonder if the approaches described in these 2 papers would be more appropriate for the existing and future whooping crane data since the cranes appear to be territorial on the wintering grounds?

   Chandler, R.B. and J.A. Royle. Spatially explicit models for inference about density in unmarked or partially marked populations. Annals of Applied Statistics 7:936-954.

   Sollman et al. 2013. Using multiple data sources provides density estimates for endangered Florida panther. Journal of Applied Ecology DOI: 10.1111/1365-2664.12098

   Particularly with a partially marked population, you could employ these methods across all the historic data to produce corrected density estimates (assuming the field maps/data are still available?). You could also then quantify the level of bias in past estimates and determine the amount of effort needed going forward. A spatially explicit approach to density estimation may also

be useful for jointly addressing your habitat use objectives (e.g., Royle et al. 2013. Spatial capture-recapture models for jointly estimating population density and landscape connectivity. Ecology 94:287-294.) These are just suggestions. If there are reasons that distance sampling would be more appropriate or efficient than these newer methods, please provide additional explanation.

Detecting and identifying color marks on whooping cranes from aerial surveys is a dangerous proposition requiring low level flights at low airspeeds (i.e., a landing-type approach; Stehn 2001). Mark-resight based techniques are not feasible from the fixed-wing aircraft platform. Ground or boat based approaches might be possible. However, undue disturbance from a close ground-level approach might bring undue stress on this highly endangered species.

Finally, the presumption of territoriality does not induce some special ability to identify unmarked individuals. We will not assume we have the ability to identify unmarked birds as was done by the previous observers. Also, unfortunately, little of the raw data from the traditional technique is still available (hence the need to standardize the technique and provide for data archiving as this protocol does).

3. Although a distance sampling approach would theoretically work, I am unsure that it would be much of a logistical improvement given the level of precision required to meet your stated objectives, the small population size, and the level of survey effort that would be required.

   We are unsure why this Reviewer suggests that distance sampling would not be much of an improvement over the traditional technique. We have provided many reasons why the traditional technique was deficient including it was unstandardized, likely could not be repeated by new observers, and lacked formal measures of uncertainty. The new technique provides a repeatable, standardized method that allows for rigorous statistical analysis. Distance sampling is very appropriate for a transect-based aerial survey and frankly, we do not understand why such techniques were not implemented earlier. As we note in the protocol, many robust techniques "have been available for decades."

   If I understand your power analyses correctly, it appears that you would need to conduct 15-20 surveys per year to meet your objective of detecting a 10-15% decline over 3 years? Is that level of effort reasonable?

   As for the power analysis and justification of the objectives, we did not provide enough detail. Thank you for catching that. Butler et al. (2013) identified scenarios that would significantly reduce abundance from the current trajectory and delay reaching the downlisting goal of 400 birds by ≥5 years. Those scenarios were ≥3 consecutive years of –9.5% annual decline or ≥2 consecutive years of –14% annual decline. To estimate one growth rate, 2 years of abundance estimates are needed. Therefore, we need to be able to detected a –9.5% annual decline over a 4-year period or a –14% annual decline over a 3-year period. We have now provided this information in the new section, "Justification of objectives" in Element 1. We have also reworded the objective to better reflect this. "Provide an estimate of whooping crane abundance within the surveyed area (see Element 3, Sampling frame) with enough precision to detect a 10–15% annual population decline over a 3- to 4-year period (see justification provided in Butler et al. 2013)."

   To detect a –15% annual decline over a 3-year period, a CV of 6% is needed and to detect a –10% annual decline over a 4-year period, a CV of 9% is needed. Based on the winter 2010–2011 pilot

work, we estimated 6 surveys would result in a CV of 7%. Based on the surveys conducted during January 2012, we estimated 5 to 6 surveys would result in a CV of 9% and 10 to 13 would result in a CV of 6%. We show a range in Table A1 because the number of surveys needed to reach a given CV depends on whether data are analyzed using HDS or CDS (see new note in Table A1). We have also updated Table A1 with additional effort estimates. We are recommending that at least 6 surveys be conducted (see "Element 3, Survey repetition and sample size" section); this does not mean more surveys cannot be conducted. However, our experience suggests that 7 or 8 surveys per year is about the logistical limit that can be accomplished in the 2-week window.

Note, that Butler et al. (2013) recommended being able to detect a –15% annual decline over a 2-year period or a –10% annual decline over a 3-year period to allow enough time to respond to the scenarios. This would require 4% to 6% CV which can be easily obtained by pooling data from multiple years as this Reviewer recommended in comment #20.

However, realize that the scenarios outlined in Butler et al. (2013) have only occurred once in 73 years of monitoring this population. This population has grown at an average of 3.9% per year since winter 1938–1939, and has exhibited a 10-year cycle in population growth (Boyce and Miller 1985, Boyce 1986, Butler et al. 2013). Therefore, a meaningful alternative to detecting the scenarios outlined in Butler et al. (2013) might be to detect a 3–5% annual decline over a 10-year period.

It also seems that because you are covering the entire area, you would not be able to address temporary emigration, correct? If I misunderstood these issues, please clarify.

The Reviewer does misunderstand the appropriate use of the gdistsamp function in unmarked. As long as complete detection on the line occurs, temporary emigration can be estimated. If temporary emigration is not occurring, then incomplete detection on the line can be accounted for. However, if both temporary emigration and incomplete detection are occurring then the parameter ϕ is confounded and uninterpretable. We discuss this issue in "Appendix C, A note about generalized HDS" and we address additional components of this Reviewer's concerns about temporary emigration below.

4. Overall, it seems like it would be worthwhile exploring the potential for utilizing both the mark-resight information and the spatial location/territory information to adjust past estimates. Going forward, I would expect that any method should utilize spatial location, knowledge of territories, and mark-resight of banded individuals to improve estimates. Distance sampling could be used as one component of this process if the probability of detecting a bird given it was present and available was of particular interest, but it seems the gains in precision may be much greater if the rest of the available information were incorporated as well.

   We will not assume we have the ability to identify unmarked birds as was done by previous observers. Some of the birds are marked in the population but we will not attempt risky flight procedures in an attempt to identify those marks from the air. We like the idea of using mark-resight data from the traditional survey in an attempt to improve historic estimates. However, much of that information is not available anymore.

5. In general the protocol reads well, although some key information is only available in the appendices. More detail on the relationship between the stated objectives and the ability of the proposed methods to meet them is needed. In my reading the stated objectives required much

more intensive sampling than the recommended ~6 surveys in order to detect a 10-15% change over 3 years. Based on the power analyses, sampling would need to be much more intensive. Please clarify how the sampling effort will meet the objectives as stated.

It would have been helpful if this Reviewer identified what key information is only available in the appendices. Without specifics, addressing this is difficult. We have provided a new section justifying the objectives and have inserted additional details throughout the protocol where needed based on the other Reviewers' recommendations. We have provided clarification of the power issue in response to this Reviewer's comment #3.

Specific comments:

<u>Element 1: Background and Objectives</u>

1. Fig. 1: font map is too small.

   We have made all map-based figures larger (e.g., Figs. 1, 5, 6, and 8). Originally these figures were only 4.5 inches wide but now they are 6.5 inches wide. We are unable to increase the font size associated with the background layers within these figures because those fonts are embedded as part of a basemap in ArcGIS. The font sizes displayed in the larger figures will have to suffice. After increasing the size of the map-based figures, we decided to increase the size of all the figures in the protocol.

2. Sampling Objective 1: Is this reasonable? Based on Appendix A it looks like you would need to conduct >20 surveys per year to meet your objective. It seems this is as many or more flights than have been conducted in the past.

   We have provided clarification of the power issue in response to this Reviewer's comment #3. During the traditional technique, the number of surveys conducted between December 1st and March 31st ranged from 4 to 21 (mean = 12.52, SD = 4.288; Butler et al. in prep). Again, another reason why the critique of the traditional technique is needed.

<u>Element 2: Pilot Studies</u>

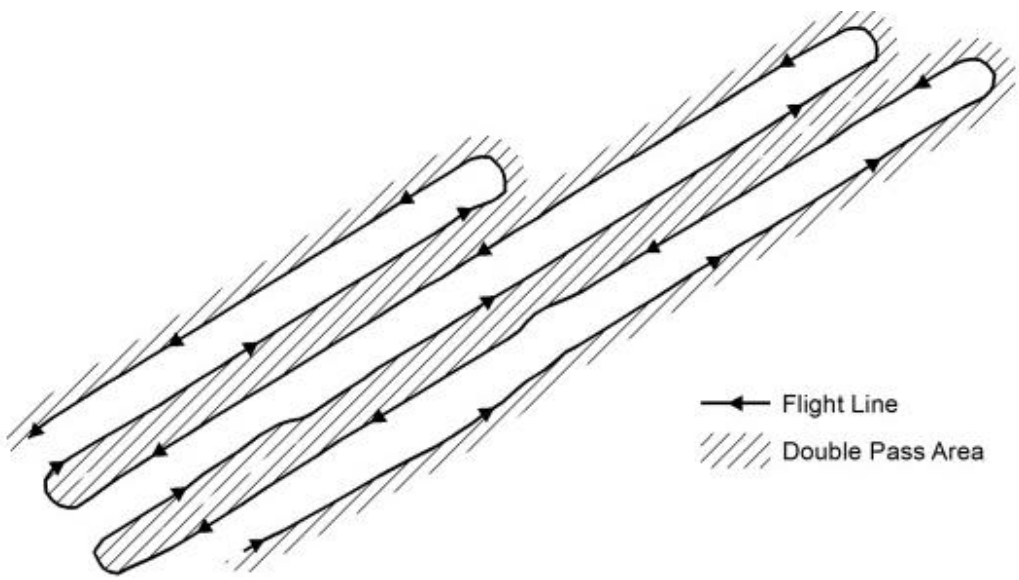
3. It may be important to require a specific type and configuration of aircraft for these surveys. As alluded to, the Kodiak has poor visibility in some areas and travels at a much higher rate of speed than planes used in the past. Tandem aircraft generally have better visibility for the observer and fly at slower speeds. Have you considered tandems (e.g., SuperCub, Husky) or are they not available?

   We have considered other aircraft but none of them are available. We have added the following sentence to "Element 4, Aircraft type." "We recommend using an aircraft similar to a Cessna 206." However, we recognize that logistics and aircraft availability may limit which aircraft is used during the survey. This is unfortunate, but often out of our control.

4. In the past surveys it appears that the pilot and observer surveyed out opposite sides of the plane correct? If so, if transect spacing was 400-500m (Stehn and Taylor 2008), then each 'strip' was covered twice, once by the pilot (partially) on the first transect and then again by the observer on the next pass. Under this scenario, detection would be much higher than that of your pilot work.

Our pilot work replicated the traditional technique (last year of surveys conducted using the traditional technique provided the data to estimate detection probabilities from the traditional technique).

The pilot did not participate as an observer but did occasionally detect whooping cranes. The observer attempted to survey both sides of the aircraft but primarily observed only one side (left-side). Stehn and Taylor (2008) reported that transects were spaced 400 to 500 m apart. However, GPS track data from the 2010 surveys indicated transect spacing varied between 250 to 800 m. Unfortunately, no GPS track data were saved by observers during the previous years. The traditional surveys were much less systematic than portrayed in Stehn and Taylor (2008) and were more targeted to areas known to be occupied by whooping cranes. We believe this is a good example of why we needed to critique the traditional survey technique. Further, as we show in the pilot work (Strobel and Butler 2014), detectability within 500 m of the survey line was only 0.558 (SE = 0.031). One might think detection was higher during the traditional survey because of the so called "double pass" by the observer. However, some areas would get a "double pass" while other areas would not get observed at all (see diagram below, assumes one observer on left-side of aircraft).



5. The high error rate for distance categories suggests that GPS locations may be required for each group. You mention this as an option, but I think the language should be stronger in 'lessons learned #4'.

   This is not in "Lessons learned #4" but #3; we strengthened the language. In "Element 4," we specified a technique to avoid having to categorize distances into bins.

6. Is there a reason that the plane does not leave the transect line to mark and count each group? It seems like this would reduce or eliminate both measurement and counting errors. I see later this will be corrected by the heads-up display?

   Deviating from the transect line to mark and count each group would reduce potential measurement and counting errors but it would also increase the duration of the survey substantially. If it requires 1 deviation from the transect for each group and the deviation is a circle

with a diameter similar to transect width (1,000 m) each deviation would add >3 kilometers. If 100 groups were detected it would increase the survey duration by approximately 2.25 hours to the survey. The value of reducing measurement and counting errors would need to be weighed against the cost of observer fatigue and safety. Further, whooping crane group density is high enough in many parts of the primary sampling frame that diversion from the transect to flyover each detected groups might cause other groups to be missed. However, occasionally observer do divert the plane from the transect to confirm a detection if necessary.

7. I disagree somewhat with lesson 6. Your pilot work demonstrated that detection is likely <1 in a single pass survey, not that it is an appropriate technique in this situation.

   Our pilot work specifically demonstrated that the traditional technique had a detection rate within 500 m of the survey line of 0.558 (SE = 0.031). It seems that this Reviewer misunderstands the traditional technique and is misrepresenting it as a double-pass survey. As shown above (see comment #4), it is not.

   Yes, our pilot work does show distance sampling is an appropriate technique in this situation. We do not, however, exclude the possibility of other techniques with our pilot work. Please take a look at Strobel and Butler (2014).

   The way I read the results of your power analyses, you'd need quite a few surveys to reach your objectives. Is that level of effort reasonable and sustainable as the population expands?

   See response to this Reviewer's general comment #2 for discussion of the power issue. We can meet precision objectives with 6 surveys per winter.

   Would spatial approaches provide more information and increases in precision?

   Yes, spatial approaches do provide more information and increases in precision, that is why we are using the spatially-explicit hierarchical distance sampling methods.

   Could you combine distance sampling with the spatial and mark-resight information to increase precision?

   Focusing on mark-resight techniques will not be productive since so little of the population was marked in most years. Further, detecting and identifying color marks on whooping cranes from aerial surveys is a dangerous proposition requiring low level flights at low airspeeds (e.g., a landing-type approach; Stehn 2001).

<u>Element 3: Sampling Design</u>

8. The plan for expanding the sampling frame seems appropriate.

   We agree that a plan for adding new areas to the secondary sampling frame is needed. However, it is not needed immediately. This population is only growing at about 3.9% per year. It will likely be many years before many of the secondary sampling frames are promoted into primary sampling frames. Only then will we need to determine additional areas for secondary sampling frames. As

we allude to in the protocol, the spatially-explicit maps of abundance can be used in the future to identify new secondary sampling frames.

9. Page 16, lines 1-7: Your objectives state that you need to be able to detect a 10-15% decline over 3 years, but that is not what is implied here. The number of required repeat surveys increases dramatically at CVs below 0.10, and it appears that 20 surveys may be required to meet your stated objectives. Please clarify you recommended sample size here or modify the objectives above.

   We have provided clarification of the power issue in response to this Reviewer's general comment #3.

10. Group size estimator: I am not certain how *unmarked* models group size, but it might be worth assessing the fit of the group size model to the observed data. In some cases the Poisson model does not represent the observed data well and can cause fairly substantial bias in estimates. It may be worth exploring simpler models for group size (e.g., individuals vs. groups).

    We are not modeling group size. As outlined in Appendix C, we use hierarchical distance sampling to estimate the number of groups in the population. We then test for size-biased detection. If size-biased detection is apparent, we adjust mean group size accordingly. We then multiply group abundance by mean group size (or adjusted group size if needed) to estimate total abundance. This is standard practice for conventional distance sampling.

11. Page 18, lines 6-15: Another alternative would be to left truncate the data to account for the partially observable strip under the aircraft if there is continued evidence of <100% detection on the line. This effectively moves the 'line' to the tire of the aircraft and as long as all groups are seen close to the plane, no bias should occur. This would help deal with individuals missed that were under the plane. It would also allow the front seat observer to search more intensively over the rest of the transect.

    Generally the idea of left truncation does not solve this problem because detectability is usually not 100% at 20 or so meters out from the line (the half-width of the band blocked by the aircraft). We have found with forward observation, detection on the line is possible in most Cessna aircraft. Further, distance sampling does not require intensive searching over the rest of the area, it just requires 100% detection on the line. Therefore, changing the technique to increase detection at greater distances from the aircraft does little to improve estimates.

<u>Element 4: Field Methods</u>

12. Have you considered wiring a cigarette plug into the aircraft? We have had good luck using small inverters to power laptops in small planes, eliminating the concerns about battery life. All our DOI planes have them, as well as our contractors, so I don't think permitting would be an issue.

    We do use a small inverter in the aircraft. Battery issues still sometimes occur because of mistakes by observers or other reasons. We likely have corrected this issue by changing to ipad-like tablets for data collection.

Element 5: Data Management

13. The data management and backup procedures seem appropriate.

    Thank you.

Element 6: Data Analysis

14. Pages 28-29: I'm not sure you can use the repeated surveys to estimate temporary emigration in the way that Chandler et al. (2011) did because groups are likely move among adjacent sample units between surveys. Chandler et al. (2011) spaced their points so that individuals wouldn't be detected on 2 different points during repeated surveys allowing separate estimation of the two detection components. In your situation, I think this potential for movement between units between surveys would cause positive bias even if detection on the line was 1.0 because you essentially have a full coverage survey. Under that scenario, I think you'd be estimating something like the superpopulation or the amount of 'use' each sample unit gets. If I am mistaken, please clarify/discuss.

    We find nowhere in Chandler et al. (2011) that they spaced survey points so that individuals would not be detected on different points during different survey reps. Movement to different cells between surveys does not cause any problems with the generalized HDS model. In fact, the point of the generalized HDS model is to characterize that movement through the estimation of a temporary emigration parameter (ϕ). Regardless of the scenario (i.e., spacing of the survey locations), abundance from this model will be biased high, as demonstrated by Chandler et al.'s (2011) equation 4, unless it is adjusted by ϕ. Following the ideas outlined in the simulation study of Chandler et al. (2011), we show that a so called "full-coverage survey" such as ours can be used to generate unbiased estimates of abundance using the generalized HDS model of Chandler et al. (2011).

    We simulated a landscape of 100 cells that were 100 x 100 m. We divided the landscape into to 2 habitat types with each habitat type being assigned to 50 cells. The first habitat type was assigned an average abundance of 2 (based on Poisson distribution, $\lambda = 2.0$) and the second habitat type was assigned an average abundance of 0.5 (based on Poisson distribution, $\lambda = 0.5$). For each iteration of the simulation, we selected the abundance for each cell based on these Poisson distributions. We also randomized the arrangement of the habitat types across the landscape for each iteration. During each iteration, the home range center of each animal in a cell was randomly selected. Then each animal was allowed to move between surveys (movement was dictated by a multivariate normal distribution allowing animals to move in and out of their cells of origin; variance-covariance matrix = [500, 0; 0, 500]). We placed a "full coverage survey design" over our simulated landscape which resulted in 10, 100-m wide transects that were 1,000 m long. For each iteration of the simulation, we conducted 6 replicate surveys between which animals were allowed to move. For each animal location, we measured its distance to the transect line and estimated its detection probability assuming a half-normal detection curve ($\sigma = 30$). We simulated the detection process with a binomial distribution (where p was the detection probability). We conducted 1,000 iteration of the simulation and estimated 95% CIs as the 2.5 and 97.5 percentiles of the simulated distribution. We have proved the R code we used below (see attached Appendix).

The landscape we simulated had a true population size of 125 animals. The generalized HDS model estimated abundance as 200.1 animals (95% CI = 159.0–247.8) but once it was adjusted for temporary emigration (ϕ), then estimated abundance was 130.9 (95% CI = 104.5–161.0). The adjusted abundance estimate was unbiased (mean =–3.9%; 95% CI = –25.6% to 19.5%). Therefore, to correctly use the generalized HDS model, abundance must be multiplied by probability of temporary emigration (ϕ). In the protocol, we already discussed this issue in "Appendix C, A note about generalized HDS."

As mentioned above, I think you could deal with the partially observed strip beneath the plane through left truncation.

Generally the idea of left truncation does not solve the problem of incomplete detection on the line because detectability is usually not 100% at 20 or so meters out from the line (the half-width of the band blocked by the aircraft). We have found with forward observation, detection on the line is possible in most Cessna aircraft. This was only an issue in the Kodiak with pontoons.

In general, I think the density estimates from each survey will be interpretable on their own and could be thought of as analogous to the historic counts adjusted for incomplete detection. It is somewhat unclear to me how you will be using the multiple surveys/estimates other than maybe considering the highest one as the max population size estimate?

The data will be pooled across the surveys from the 28 November–26 December. This will result in one estimate of abundance, HY:AHY ratio, number of adult pairs, and number of recruitive pairs, and one spatially-explicit model of abundance. Though we believe this is clearly articulated in the methods outlined in Appendix A and C, we add a note to "Element 6, Analysis methods" section. Realize, the Elements are really just summaries and the details abound in the SOPs and Appendices. We are a little disappointed this reviewer did not examine Appendix C closer.

(I see that many of these issues are addressed in Appendix C, page 97, but I think it should be made more clear in the main text that the temporary emigration piece will probably not be usable under the proposed design.)

We have added the following statement to "Element 6, Data analysis" section. "Currently, we plan to pool all the surveys from the 28 November and 26 December survey period and use the model configuration that ignores temporary emigration (see Appendix C–A note about generalized HDS)."

15. For clarity, abundance estimates for each survey will represent the portion of the population using the survey area the day of the survey (i.e., birds temporarily located in areas outside the primary sampling units are not accounted for).

    For clarity, all surveys are pooled for analysis. We emphasize that our estimates will only pertain to the primary sampling frame in multiple places throughout the protocol (e.g., "Element 1, Objectives" and "Element 7, Summary statistics of interest"). We limit the survey to a 2-week window in order to provide as much population closure as possible.

16. How are you accounting for the spatial autocorrelation induced by your small (1 km) subunits? When all are sampled sequentially, there will be a fair bit of autocorrelation present. Will this information only be used for the habitat component?

Spatial autocorrelation or lack thereof has nothing to do with the size of the subunits. Spatial correlation could occur regardless of the size of the subunits. Spatial correlation could affect the standard errors of the parameter estimates but the magnitude of effect would be minimal. Thus, the inferences gained from our models about the effects of habitat on whooping crane abundance would not change if potential spatial correlation was accounted for. Further, there is not currently a tractable way to deal with spatial correlation in these already complex models.

Element 7: Reporting

17. The reporting requirements seem appropriate. Requiring at least an annual report that is placed in a government series is a good idea in my opinion. One suggestion would be to include periodic publication of results in the peer-reviewed literature (e.g., every 5 years). Over time this would establish your methods and results as scientifically defensible and would help support future delisting decisions.

    Thank you, we agree annual reporting is important and have provided for it in "Element 7: Reporting." We agree that periodic publication is important and we will be pursuing publication of our results. We added the following sentence to "Element 7, Reports" section. "Also, we recommend periodic publication of survey results in the peer-reviewed scientific literature."

Element 8: Personnel Requirements and Training

18. These seem appropriate.

    Thank you.

Element 9: Operational Requirements

19. These seem appropriate.

    Thank you.

Appendix A: Power Analysis

20. These analyses seem appropriate given the limitations of TRENDS and the basic assumptions made. Multiple sources of heterogeneity could reduce the power to detect trends, but the approach used is reasonable. Have you considered the increase in power that might be possible either through analyzing multiple years' data together or using a Bayesian approach with informed priors on the detection function? That should increase your precision over time assuming the detection process is similar through time.

    Yes we have considered this and will likely implement some sort of data pooling in the future to bolster our power. Thank you for the suggestions.

Appendix B: Indexing Recruitment

21. As long as this is treated as an index rather than abundance the methods seem appropriate.

    Agreed, thank you. However, we are not talking about abundance but a ratio.

Appendix C: Data Analysis

22. Page 97: As described above, I don't think temporary emigration is estimable under your current full-coverage design. If you disagree, please explain.

    See our responses to other comments regarding this issue (see Reviewer #2, comment 14). Temporary emigration is estimable under a so called "full-coverage design."

REVIEWER #3:

It is clear that a considerable amount of work went into this protocol. The techniques and analyses used within this protocol seem appropriate and logical.

Thank you. Indeed much work has gone into developing this protocol.

This is a sound protocol but needs, in my opinion, more details with explicit statements and discussion. You make many implied statements or cover specific topics but they need to be brought together in a clear and concise manner. I make specific comments below.

We thank you and the other Reviewers for their comments. Those comments have allowed us to fill in details in import places throughout the protocol.

Specific comments:

1. There are several grammatical edits needed throughout the document. I will provide two examples:

   A) Element 1, Page 1, Lines 5-6: The focus is the crane and not the "endangered species." I suggest the following change.

      Whooping cranes (*Grus americana*), an endangered species, declined to near extinction by 1941 (Canadian Wildlife Service [CWS] and U.S. Fish and Wildlife Service [USFWS] 2007).

      Agreed, changed as recommended.

   B) Element 1, Page 1, Line 10: Delete the "s" in "…..breed~~s~~ on and around Wood Buffalo National Park ….. ."

      Agreed, changed as recommended.

2. The authors make implied statements in the document on how their study design and pilot studies demonstrate how the surveys will meet the assumptions of Distance Sampling. It is my opinion that you need a section dedicated to that information. You must explicitly state how your study design and pilot studies have shown or how you will meet those assumptions. A good place may be in Element 3.

   We agree that assumptions must be addressed and we did address them. "Element 3, Sources of biases" section explicitly addressed the assumptions of our techniques.

3. You need to make a statement as to why the plane does not alter the movement or behavior of the cranes. It may be obvious to you but not to others.

   We added the following paragraph to "Element 3, Sources of biases" section:
   "Movement of whooping crane groups in response to the aircraft prior to detection could bias estimates of abundance. Movement that is independent of the aircraft causes no problems and movement after detection is not a problem as long as the initial location can be accurately determined (Buckland et al. 2001). Movement towards the aircraft by whooping cranes would result in positive bias but evasive movements would result in abundance being biased low (Buckland et al. 2001). However, attraction to or avoidance of the aircraft by whooping cranes is not an issue. During previous surveys, we observed the immediate response of hundreds of whooping cranes groups to the aircraft. We did not observe any whooping cranes being flushed into flight or make movements of >100 meters in response to the passage of the aircraft. Further, initial locations of whooping crane were easily established. When visibility allowed the detection of groups from the next transect, groups were often detected in the same location, suggesting movement away from the aircraft was not substantial."

4. You need to discuss randomness of line transect placement. Having a random start point or time is great to account for temporal bias but does nothing for spatial bias. Random transects provide unbiased coverage of landscape heterogeneity and distribution of focal wildlife species. I understand that you may survey the entire area but you need to explicitly state how that justifies the nonrandom placement of transects. It is even more important to explain given your statement on Page 12…."However, as more regions are added to the PSF as the whooping crane wintering grounds 46 expand, complete sampling will likely become infeasible." You then must justify, with less than full coverage of the study area, why you did not use randomly placed transects.

   We agree that sampling strategy is a very important consideration influencing the rigor of any survey. Hedley and Buckland (2004) indicate that spatial distance analyses such as those we use are not as dependent upon random sampling because environmental covariates are incorporated directly into the model and thus do not need to be controlled through randomization. This allowed us to place transects anywhere we wanted across the survey blocks because the environmental covariates are used to control any potential bias. We chose to place transects so that coverage of each survey block was maximized. Transects were systematically placed 1,000 meters apart.

   "Spatial modelling does not require that the lines or points are located according to a formal survey sampling scheme" (Hedley et al. 2004:48).

5. I assume that "groups" are "clusters?" Per the Distance Sampling lingo. You might want to make a reference early in the document that groups are your clusters.

In our experience most readers are more confused by the term cluster than they are by group. Usually when cluster is used it is followed by parenthetic phrase like "i.e., a group of birds." We added a note in "Element 3, Sampling units" that groups is referring to the cluster lingo used in Buckland et al. (2001).

6. I would make a quick statement somewhere as to why it is important to measure perpendicular distance to the center of the cluster or group. The statement on page 52 is not enough.

   We are glad you pointed this out since the protocol does not specify that distances need to be measured to the center of the group (except in SOP1). We have now added information regarding measuring distance to the group center to "Element 3, Sources of biases" and "Element 4, Field data collection methods." As for explaining why, SOP 1 is not the appropriate place to explain why since it is a step-by-step explanation of what the observers are to do.

7. It is important to <u>explicitly state and discuss</u> what will be considered a group or cluster and what are the criteria (e.g., behavior, separation distance, etc.) used to separate those groups. I know you make implied statements, "a group is greater than 1 crane" and "any crane within 100 m is counted within the group," but that is not enough. You must justify that criterion with data or citations. This will strengthen the value of your abundance estimate.

   Groups of whooping cranes are primarily 1, 2, 3, or 4 individuals (usually a family group) and the occasional larger group of subadults. We added this statement to "Element 3, Sampling units" section.

8. Determining size bias is not the only important item when it comes to group size. A consistent size from group to group will improve or strengthen your CV. A consistent size from group to group can be controlled sometimes with a good, a priori, definition of what a group is and a procedure to separate groups.

   Groups of whooping cranes are primarily 1, 2, 3, or 4 individuals (usually a family group) and the occasional larger group of subadults. It is not possible to control group size; it is what it is. However, group size is a minor part of the total variance of abundance estimates from these surveys and, therefore, has a minor impact on the CV.

9. Being a little facetious, but is a survey the act of traversing a transect or is it the amalgamation of all transects being flown. The author needs to <u>explicitly state</u> what a survey is. The use of the term survey is loosely used throughout the document. The description/definition of a survey, segment, and transect need to be tightened up. Segments make up transects and transects make up a survey. A completed survey is when all segments of all transects are flown. You mentioned at least two surveys will be flown during peak times. The explicit definition of what you are calling a survey is vital here.

   A survey is all transects within the sampling frame under discussion. We provided clarification throughout the protocol that 6 surveys of the PSF means that all transects will be surveyed 6 times and that 2 surveys of the SSF means that all transects will be surveyed 2 times.

10. Suggestion: I would fly each transect at least twice (on separate days, preferably back to back) and define that as 1 survey. If you fly each transect 4 times (4 separate days) then you would have 2

complete surveys. By flying each transect (alternate directions) twice per survey, you strengthen the abundance estimate and CIs by attempting to capture the inherent variation (e.g., crane behavior, environmental factors, etc.) that is present.

Thank you for the suggestion. We are flying the survey (all transects in the primary sampling frame) at least 6 times. We never start a survey with the same survey block. Transects are usually flown in different directions than the last time.

11. Please don't combine objectives when providing information, such as on Page 14/Line 20. Discuss objectives individually.

    When discussing the reasons for selection of survey timing or other protocol constraints, we combined objectives when the reason was similar. This is simply more efficient and reduces the overall length of the protocol. Thank you for your suggestion but we opted for efficiency here.

12. Regardless of plotting locations on a map within a GIS, I imagine you will still need to bin your perpendicular distances. Perpendicular distances must be precise and human error will still be very much present with hand plotting. I would discuss that binning will still be part of the data analysis as exact distances can't be made.

    Beyond binning because of potential errors associated with distance measurements, we must bin our data for analyses because the hierarchical distance sampling model we are using is the multinomial-Poisson mixture model of Royle et al. (2004). This multinomial part of this model requires data to be binned. Binning does not necessarily fix problems with measurement error.

13. Will binning be used to create the all imperative shoulder on the detection probability histogram. You should probably discuss how the data will be handled if an apparent shoulder is not found in the data.

    Our data have exhibited a shoulder in the past for this survey and there is no reason to expect it not to in the future. Therefore, we chose not to discuss this issue in the protocol. If users are interested in this issue, we suggest they consult Buckland et al. (2001).

14. Element 1, Page 1, Lines 26-27: This sentence is confusing. Are you monitoring the metrics or are you stating the quantitative population metrics from monitoring are important for achieving criteria? I suggest rewording this sentence to be clearer.

    The sentence in question states, "Monitoring quantitative population metrics is important for determining if the population has achieved downlisting criteria." We do not see how this is confusing. The population must be monitored in order to know if the downlisting criteria have been achieved. We choose not to change the sentence since the Reviewer provided no suggestions on how to improve or clarify the sentence.

15. Element 1, Page 2, Line 1: I suggest the following addition "…..ensure population closure (e.g., emigration, immigration; Lancia et al. 2005, Morrison et al. 2008, Stehn and Taylor 2008, Conroy and Carroll 2009)."

    Agreed, changed as recommended.

Element 1, Page 2, Lines 2-5: I suggest the following additions "…..individuals is usually less than 100% (i.e., <1) and 2 detectability is influenced…."

Agreed, changed as recommended.

Element 1, Page 2, Lines 9-10: I suggest the following modification "Therefore, a~~n~~ robust alternative method that does not…."

Agreed, changed as recommended.

16. Element 1, Page 3, Lines 3-4: I suggest the following addition, "….location on a paper copy of a 1:46,080 Digital Orthophoto Quarter Quadrangle (DOQQ)."

    We added "Digital Orthophoto Quarter Quadrangle" as suggested.

17. Element 3, Page 9, Line 27: I would suggest deleting this secondary heading, "Sampling Design."

    Element 3, Page 9, Line 35: This is a very odd heading. I would suggest changing the heading from "Target universe" to "Target inference" or something a little more descriptive.

    Thank you for your recommendation. However, we choose not to change these headings since they are part of the I&M Protocol Template.

INTERNAL REVIEWER #1:

First of all, many thanks to you and your staff for undertaking the monumental task of getting a protocol in place. I know first-hand how time consuming and difficult this was, particularly with ongoing staffing changes, etc. I think the protocol puts us in an excellent position moving forward as we evaluate metrics needed for species recovery, Refuge acquisition decisions, etc. I am excited about the upcoming season and the opportunity that the well-thought out protocol and additional staffing capacity provides us, ensuring that we get things "right". Attached are some of my initial thoughts and comments, I wanted to get you something as quickly as possible so we can continue the upcoming season's planning efforts. Let me know if you have questions about any of my comments.

Thank you for your thoughtful review; we value your comments.

Specific comments:

1. Pg. 1, line 37: remove "the"

   Agreed, changed as recommended.

2. Pg. 2, line 3: change "influence" to "influenced"

   Agreed, changed as recommended.

3. Pg. 4, line 15: consider adding “reliably” to “aerial surveys cannot be used to reliably imply mortality.”

   Agreed, changed as recommended.

   It may be worth pointing out that there were marked individuals during a portion of the 1950-2010 surveys and the marked individual’s presence and location was used to establish observer knowledge of territoriality, etc. For the most part, this is no longer the case and detection of marked individuals from an aerial survey is nearly impossible given the protocol for flight speed and altitude.

   We added additional discussion of the potential problems associated with the traditional estimates of winter mortality. We also added a note about territoriality and site fidelity were derived from limited observation data from marked birds.

   Wild whooping cranes were not individually marked until 1975. In fact, only 132 marked birds have been in the population spanning the period of 1977 through 2004 (Gil-Weir et al. 2012). Over 40% of the years in which the traditional survey was conducted (1950–2011), no color marked birds were in the population. By the winter of 2010–2011, only 7.8% of the flock was marked (Stehn 2011). It is also important to note that many of the banded cranes were not marked with uniquely identifiable color marks but with aluminum bands (Stehn 2011).

4. Pg. 4, line 36: Consider stating that “mark-recapture models using information from marked whooping cranes to estimate mortality may be explored in future.”

   We are unsure mark-recapture based estimates of mortality are possible given the difficulty of detecting marks during the aerial survey. However, we did add the statement, “mortality estimates are better derived from radio- or GPS-telemetry based monitoring of a sample of the population.”

5. Pg. 5, line 20: Suggest conducting surveys annually until this precision goal is achieved over 3 yrs.

   We recommended annual monitoring in “Element 3, Monitoring frequency” section. But we discuss the potential of biennial monitoring which could meet the survey’s objectives if the precision of surveys is adequate.

6. Pg. 7, line 17: Might want to emphasize that the decoy study was conducted when WHCRs were not present at Aransas, eliminating confusion with live birds.

   Agreed. We added the sentence, “the experiment was conducted during September prior to the arrival of whooping cranes to eliminate potential confusion with live birds” for clarification.

7. Pg. 10, line 20: add “primary” to “apply to the primary sampling frame”

   Good catch; changed as recommended.

8. Pg. 12, line 20: Should consider some type of density estimate to serve as a threshold for including SSF into PSF. The way it is structured now, relatively small SSFs (i.e., Holiday Beach) have the same threshold applied for inclusion into PSF as those that are much larger (i.e., Guadalupe Delta). Given

this fact, some small SSFs may never be included into the PSF even after several years of consistent WHCR use. Consider analyzing what the average WHCR density is in the PSF and using that to determine an inclusion threshold.

Great idea! We have now estimated a threshold of number of detections for promotion of each SSF to PSF based on minimum density estimated in the PSF regions. New text in protocol is:

"Previously collected data indicated that whooping cranes have sporadically used several areas outside of the PSF. The SSF will be used to monitor expansion of the whooping crane wintering grounds. We determined a minimum number of whooping cranes groups that must be detected in a SSF region, before that SSF region will be included in the PSF for future surveys. The minimum number of groups (G) for SSF region i was determined based on the following formula:

$$G_i = \text{floor}(D \cdot A_i \cdot p)$$

where D was the minimum PSF region-specific density of whooping crane groups estimated during the winter 2013–2014 surveys (≈0.1 groups/km^2), A_i was the area of SSF region i, and p was the average detection probability during the 2013 surveys (0.7 detection probability within 500 m of transect). We rounded each G_i down to the nearest integer. The estimated threshold (G_i) of groups for a SSF region must be detected $\geq$2 times within a survey year before that SSF region will be promoted into the PSF for future surveys (beginning the next year). The minimum number of whooping crane groups needed to promote a SSF region into a PSF region varied from 1 to 9 (Table 2)."

We also add a new table to summarize the new thresholds.

9. Pg. 13, line 25: May want to note that surveys in WBNP are primarily used to estimate WHCR nests and fledged chicks, not abundance. I can see what protocols CWS has if you would like.

   Please check on those protocols, if any exist, and let us know how CWS's surveys are currently conducted and for what purposes. We believe integration of information from the abundance surveys on the wintering grounds with information about reproductive output on the breeding grounds could provide valuable insight into the limiting factors for this population.

10. Pg. 14, line 2: The 2-wk window may be too narrow once weather and logistical issues (plane maintenance) are considered. Consider changing to 3-wk window.

    We have decided to keep the 2-week window for now and see how things work out for winter 2013. We do not necessarily consider the 2-week window an absolute. It emphasizes the importance of getting the surveys done in as short of a window as possible. And a 2-week window is realistic. However, if it takes 16 days to complete the 6 surveys instead of 14 days, that is alright.

11. Pg. 14, line 13: Seems as though a clear objective for surveys done in secondary periods has already been established ("to document temporal changes in resource use"). May want to clarify this and note that surveys during secondary periods are optional and will be conducted only if primary objectives have been met and resources allow.

We do not consider "to document temporal changes in resource use" as a sufficient objective. During the first year of the new survey, we conducted surveys in a few different time periods during winter. Though there were slight changes in habitat use through time, the primary driver of all the models remained percent saltmarsh. We agree that surveys during secondary periods are optional and did note this in "Element 3, Survey timing" section. The primary reason for conducting the surveys is to estimate abundance, not quantify temporal changes in resource use. Models of resource use are a useful byproduct of the statistical analyses but resource use is best studied using other means, such as GPS-telemetry.

12. Pg. 15, Figure 9: Given the historical nature of this graph, I am concerned that we could be starting primary surveys too early in the year, potentially not capturing peak abundance in the early surveys. Should later migration events be accounted for?

    Based on the historic data we felt that surveys probably could begin either the second or third week of November. However, we delayed starting surveys until December just in case migration was delayed some years. We agree that if we see that migration is delayed in a particular year, then we should delay surveys as appropriate for that year. In fact, that is part of the reason we allowed surveys to be conducted anytime during the 1-month period (as long as they are blocked together into a 2-week window). If a trend of later migration dates becomes apparent, then we should re-evaluate the survey dates for future surveys.

13. Pg. 17, line 4-5: As stated previously, this threshold seems somewhat arbitrary considering the SSFs vary greatly in size and habitat quality.

    Agreed and we have addressed this issue in this Reviewer's comment #11.

14. Pg. 17, monitoring frequency: Need to explain how sampling objective 1 ("enough precision to detect a 10-15% population decline over 3 yrs") can be met without annual sampling.

    The paragraph in question is a discussion beyond objective 1. It broaches the issue current monitoring may be unnecessary until the population is closer to reaching the downlisting goal. We feel it is appropriate to discuss this point of view since it is an alternative to the objective outlined in the protocol. At current levels of precision, biennial monitoring probably could not meet current objectives. However, as the kinks are worked out in the new protocol, precision may improve.

15. Pg. 18, line 17: Is it worth saying that if an observer has any doubts as to the species ID during a survey, the bird should not be recorded? Or come up with a similar rule?

    Yes, this is worth saying. We had assumed that observers would only record detections they were certain to be whooping cranes. We have made this explicit in the protocol in the section in question and we added "if an observer has any doubt as to the species, that detection should not be recorded" to SOP1.

16. Pg. 20, line 29: Note that full Nomex fire clothing (shirt and pants) can be substituted for a flight suit per regs.

    The Reviewer is correct. See ALSE Handbook. We added a reference to the ALSE Handbook to the protocol and changed from "Nomex flight suit" to "Nomex flight suit or other fire-resistant clothing."

17. Pg. 20, line 30: I have not seen Nomex gloves identified as a PPE requirement, need to check on this. It would make data recording much more difficult.

    The Reviewer is correct; leather gloves maybe used instead of Nomex gloves. See ALSE Handbook. We changed "Nomex gloves" to "Nomex or leather gloves."

18. Pg. 20, Table 2: We should check to see if voice-activated voice recorders are available. This would reduce data loss associated with on/pause toggle issue.
    The Refuge Biologist was looking into this as a possibility; also, one observer has switched to simply writing down their observations. We encourage observers to use whatever technique works for them without compromising the survey's integrity.

19. Pg. 21, line 19: This list should be created prior to survey season and used accordingly by pilot and observers.

    Yes it should. Thanks for emphasizing the importance of making this list prior to the start of surveys. We have added this to SOP1.

20. Pg. 21, line 38: Per our experience last year, it takes about 4.5 hrs to conduct the survey (PSF). The stated time allotment doesn't leave any time to conduct SSF as suggested earlier in the document. Is it okay to conduct SSF flights outside of 10:00 to 15:00 hrs?

    We should attempt to keep all surveys inside the 10:00 to 15:00 timeframe. However, some deviance is probably acceptable.

21. Pg. 23, line 22-24: I still think we should explore a drop-down menu available via DNR Garmin or other software that would allow instantaneous data recording of group size/type. This would eliminate the need to voice record location and #/type of crane group detected and lowers potential data loss associated with this step. Each observer would still use a voice recorder for notes.

    We agree alternative data recording options need to be explored in the future. As mentioned before, we encourage observers use whatever technique works for them without compromising the survey's integrity.

22. Pg. 32, line 19-20: Need to state who is responsible for the annual report. Refuge biologist in conjunction with I&M, WHCR coordinator?

    Thank you for catching that oversight. We have added the following sentence to the "Element 7, Reports" section and the "Element 8, Roles and responsibilities" section. "Compilation of interim and annual reports will be a collaborative effort among the Lead Biologist, Regional Biometrician, Regional Data Manager, and Whooping Crane Recovery Coordinator (see Element 8–Roles and responsibilities). Update reports are the responsibility of the Lead Biologist with assistance from other observers."

23. Pg. 33, line 4-9: Is it worth stating that at least one alternate observer should be trained if possible in case one of the primary observers is not able to conduct the survey. This is particularly important given the survey's time constraints.

Good idea. We added the following to "Element 8, Training" section. "At least one alternate observer should be trained in case one of the primary observers is not able to conduct a survey."

INTERNAL REVIEWER #2:

I would like to thank the authors for giving me the opportunity to review the whooping crane abundance survey protocol. Below are my line by line comments followed by general overall comments.

Thank you for taking the time to evaluate and critique our protocol. We appreciate your efforts.

Specific Comments:

1. Cover Page: Insert "(Protocol)" after the word "Survey"

   Agreed, changed as recommended.

2. Page x, Line 11-13. Rewrite to: Stehn and Taylor (2008) briefly described the aerial survey methods employed from 1982-2011 and identified multiple factors that affect detection of whooping cranes.

   We disagree. Stehn and Taylor (2008) identified more than just factors that affect detectability. For example, they discussed problems associated with not predefining flight paths, whooping crane movements during surveys, upland use by whooping cranes, etc…

3. Page x, Line 17-18. Rewrite to: Regional and refuge staff then launched a multifaceted effort to develop a statistically rigorous survey method.

   Thank you for the suggestion, we reworded the sentence as "Regional and refuge staff then launched a multifaceted effort to improve the survey methods and develop a statistically rigorous technique."

4. Page x, Line 40. Rewrite to: The survey methods in this protocol provide a statistically rigorous means to estimate the annual peak abundance, and associated degree of confidence, of whooping cranes wintering within the sampling frame. Furthermore, the survey methods provide estimates of….

   Thank you for the suggestion, we reworded the sentence as "The survey methods provided in this protocol are a defensible, statistically rigorous means to estimate the annual abundance, and associated degree of confidence, of whooping cranes wintering within the sampling frame."

5. Page 1, Line 11. "Gulf" should be capitalized.

   We disagree. Texas gulf coast is not a proper noun. The Gulf of Mexico is a proper noun. The grammar is similar to "south Texas." The words "south Texas" does not indicate a state like West Virginia and therefore, should not be capitalized. The same applies to the Texas gulf coast.

6. Page 1, Line 28. Replace "defensible and creditable" with "statistically rigorous"

The sentence in question reads, "monitoring also provides a tool to measure recovery and bolster conservation efforts by providing defensible and creditable data that can be used to inform decisions affecting whooping crane conservation and management." This sentence is discussing monitoring in general terms, not a specific monitoring technique. We believe it would be better to exclude either phrases from this general sentence and reserve such phrases for discussing a specific monitoring technique. Therefore, we have reworded the sentence as "Monitoring also provides a tool to measure recovery and bolster conservation efforts by providing data that can be used to inform decisions affecting whooping crane conservation and management."

7. Page 2, Line 5-8. Rewrite to: Stehn and Taylor (2008) recognized these sources of potential bias and also indicated that population increases and range expansions would decrease the accuracy of the census.

   Thank you for the suggestion, we reworded the sentence as "Stehn and Taylor (2008) recognized these sources of bias but they did not address them. They also indicated that as abundance increases and the whooping crane's wintering grounds expand, the accuracy of their "census" attempts would decrease."

8. Page 2, Line 9-12. These two sentences should be moved to just above "Objectives" on Page 5.

   Thank you for your suggestion but these two sentences are appropriate in their current location.

9. Page 3. Delete blank line 7 and move Line 8 (sentence beginning "Each") up to line 6.

   We agree that this sentence should be part of the previous paragraph.

10. Page 3. Delete line 10-31 (sentence beginning The "peak population size" and ending with the graph and Figure 2 caption). This paragraph has nothing to contribute to the understanding of why the Service is changing methods. The "Traditional census estimate and Maximum count during the survey year" are clearly correlated. The use of a T-test to statistically show the minor difference between the two does not seem appropriate and to report a 3 bird difference as significant seems absurd.

    The section in question reads, "The reported winter abundance was presumably obtained from the survey with the most individuals detected but was rarely replicated in subsequent surveys. However, the survey with the most individuals detected rarely corresponded with the winter abundance reported in the International Recovery Plan (CWS and USFWS 2007) and Aransas NWR annual reports (Stehn 2009, 2010, 2011; Figure 2). In fact, the reported abundance was ≈3.7 birds (t = 3.754, df = 30, $P < 0.001$) more than the maximum count observed during the 1980–2010 surveys. This likely occurred because birds observed outside of the survey area by the public or other parties were included in the reported abundances. Interestingly, during 1980–2010, the maximum number of individuals detected during a survey was greater than the reported abundance for 6 of the survey years (Figure 2). This inconsistency may be attributed to the observer's interpretation or perception of double counts. However, we cannot know the sources of these inconsistencies since rules governing how data from separate surveys were combined or how public reports of whooping cranes from outside of the "censused" area were incorporated was largely determined by the observer's opinion. This shows the traditional "census" was not an absolute enumeration of the population but instead a relative index of abundance."

Understanding the limitations of the historic data only strengths our ability to use those data; this section is important to help readers understand those limitations. We believe the inability to discern exactly how numbers were derived is a fundamental issue associated with the traditional technique. Instead of assuming those data are absolute, we think it is better to couch our inferences within the data's limitations and avoid risking drawing conclusions that are beyond the scope of the historic data.

Discussing a 3 bird difference is not absurd. Three birds can fundamentally change the interpretation of these data since they have been treated as absolute truth. For example during the winter of 2008–2009, the traditional technique reported 270 birds on the wintering grounds with 23 of them dying during the winter. During the winter of 2009–2010, the population was reported to be 264 birds with 22 juveniles which would mean there were 242 adults in the population. However, the maximum count during the winter of 2009–2010 was 270 birds which would presumably mean there were 248 adults in the population. If 23 mortalities occurred during the winter of 2008–2009, there could only be 247 adults in the population during the winter of 2009–2010 (270–23=247; assuming no mortality during other parts of the year). Where did that extra adult come from? Did the maximum count of 270 during the winter of 2009–2010 include 6 birds that were counted twice? Was mortality during the breeding- and migration-periods only 2% ((247–242)/247)? This is an illustration of the fundamental danger of a technique that portends to deal in absolute truth when there is really much uncertainty in the estimates.

11. Page 4, line 1-7. Move up to page 3 line 6 just ahead of current line 8 on page 3.

    Since we chose not to delete the previous paragraph, these lines should remain where they are.

12. Page 4, line 7-15. Sentence beginning "Group structure" start as a new paragraph. Rewrite to: In the past, group structure and location data was also used to assist in quantifying mortality of individuals. For example, subsequent observations of a group of 1 adult and 1 chick where previous surveys located a group of 2 adults and 1 chick could result in the presumption that 1 adult had died. However, the "unusual movements and behaviors of color-banded whooping cranes" observed by Stehn (1992), such as extra-territorial excursions, rapid pair formation after mate mortality, pair-bonded individuals overwintering apart, and occasional departure from the wintering grounds during the winter period, in addition to incomplete detection of individuals makes using group structure and location data of unmarked birds from aerial surveys problematic.

    We added more information to this paragraph in order to help readers better understand the problems associated with the mortality estimates derived from the traditional technique. The section is now split into 3 paragraphs:

    "Wintering whooping cranes are thought to demonstrate territoriality and site fidelity within and across winters (based on limited observation data from marked birds; Stehn and Johnson 1987, Stehn 1992, Bonds 2000). The fidelity of whooping cranes to their wintering territory has been used in conjunction with group structure as a surrogate for individually marked birds. Groups of equal size found in similar areas on subsequent surveys were assumed to be the same individuals. These data were used to identify circumstances where distinct groups, or individuals within groups, were not detected and therefore additional search efforts were conducted (Stehn and Taylor 2008)."

"Group structure and location data was also used to assist in quantifying the mortality of individuals (Stehn and Strobel 2012, Pugesek et al. 2013). For example, subsequent observations of a group of 1 adult and 1 chick where previous surveys located a group of 2 adults and 1 chick resulted in the presumption that 1 adult had died. However, the "unusual movements and behaviors of color-banded whooping cranes" observed by Stehn (1992), such as extra-territorial excursions, rapid pair formation after mate mortality, pair-bonded individuals overwintering apart, and occasional departure from the wintering grounds during the winter period, in addition to incomplete detection of individuals indicates that group structure and location data of unmarked birds from aerial surveys cannot be used to reliably imply mortality."

"The technique for enumerating mortalities assumed that if an individual whooping crane was not observed for 2 or more surveys, that individual was dead (Stehn and Strobel 2012). For this technique to produce consistent results, the number of surveys conducted per winter needed to be relatively constant. However, the number of surveys conducted between December 1st and March 31st of each year ranged from 4 to 21 (mean = 12.52, SD = 4.288; Butler et al. in prep). This deficiency in the technique manifested itself in an inverse relationship between winter mortality estimates and the number of surveys conducted between December 1st and March 31st (binomial regression: odds ratio = 0.932, $\hat{\beta}$ = –0.071, SE = 0.028, W = 6.595, P = 0.010; Butler et al. in prep). Hence, more surveys resulted in less mortality because temporary emigrants had more chances to return and be re-observed on their territories. For example, imagine that 5 consecutive surveys were conducted and a whooping crane was missing from its territory on the third and fourth survey occasions. If the last survey had not been conducted, that whooping crane would have been considered dead."

13. Page 4, line 35. Delete entire bullet.

The bullet in question is "The observer assumed ability to uniquely identify unmarked individual whooping cranes." This is a true statement, and relevant, so we are unsure of the request for deletion. While some believe that previous observers could consistently identify unmarked whooping cranes, there are no data available to support this belief. In our professional experience, defensible identification of free-ranging individuals requires the use of unique markers. As such, we feel our comment is accurate. This point forms an added example of why we needed to critique the traditional survey technique.

14. Page 4, Line 37. Stop the sentence after the word "tested." Delete "and is likely untenable."

The statement in question is "The suitability of using group structure and territory location as a surrogate for individually marked birds to measure mortality rates has not been quantified or tested and is likely untenable." We have changed this sentence to "The suitability of using group structure and territory location as a surrogate for individually marked birds to measure mortality rates has not been tested and is likely untenable." We have provided additional information in a previous paragraph to explain the problems associated with estimating winter mortality from the traditional survey technique.

15. Page 5, line 1-3. Delete entire first two sentences

The sentences in question are "Many technological and statistical advances have resulted in improvements in data collection and analysis techniques that were not readily available when aerial

surveys began in the 1950s. However, many of these advances have been available for decades (e.g., Burnham et al. 1980)." These statements are valid and they make us wonder why previous biologists did nothing to address problems with the traditional technique when much better techniques have been available for decades. We are updating the survey, and the methods we use are tested and established.

16. Page 5, line 6. After the sentence ending "data." Add the two sentences from Page 2 Line 9-12

    We edited as suggested.

17. Page 5, line 7. Replace "rigorous, defensible" with "statistically rigorous"

    We removed the word defensible.

18. Page 5, Line 9. Add another paragraph with this text: Quantifying winter mortality will not be a stated objective of this monitoring protocol due to the multiple factors previously identified regarding group structure and location data of unmarked birds. Estimates of mortality will require an additional protocol.

    We refrained from adding this sentence because we will not be developing an additional protocol to estimate winter mortality. Winter mortality has little influence on population change but recruitment and mortality during migration have large influence on population change (Butler et al. in prep.). This protocol indexes recruitment but radio- or GPS-telemetry based monitoring are probably the only ways to better understand the causes of mortality (this is an ongoing research project).

    We did add the statement, "mortality estimates are better derived from radio- or GPS-telemetry based monitoring of a sample of the population," to the "Element 1, Traditional "census" effort" section.

19. Page 5, line 14. Replace "defensible" with "robust"

    Ok, thank you.

20. Page 5, line 17. Delete "including the delineation of critical habitat" Critical habitat has already been established and until such time that the Service formally decides to modify the critical habitat units I would refrain from stating that this will be one of the uses of the data.

    Good catch; we did not intend for this to mean "critical habitat" in the legal sense of the term. We have changed term to "important habitat."

21. Page 5, line 20-22. I think being able to statistically show a 10 to 15% decline over a three year period is exactly what we should be striving for. This would be a 3-5% decline annually. This does not seem to match the text in Butler et al. 2013 which is used as a reference here.

    We did not provide enough detail for the power analysis and justification of the objectives. Thank you for catching that. Butler et al. (2013) identified scenarios that would significantly reduce abundance from the current trajectory and delay reaching the downlisting goal of 400 birds by $\geq$5

years. Those scenarios were ≥3 consecutive years of –9.5% annual decline or ≥2 consecutive years of –14% annual decline. To estimate one growth rate, 2 years of abundance estimates are needed. Therefore, we need to be able to detected a –9.5% annual decline over a 4-year period or a –14% annual decline over a 3-year period. We have now provided this information in the new section, "Justification of objectives" in Element 1. We have also reworded the objective to better reflect this. "Provide an estimate of whooping crane abundance within the surveyed area (see Element 3, Sampling frame) with enough precision to detect a 10–15% annual population decline over a 3- to 4-year period (see justification provided in Butler et al. 2013)."

To detect a –15% annual decline over a 3-year period, a CV of 6% is needed and to detect a –10% annual decline over a 4-year period, a CV of 9% is needed. Based on the winter 2010–2011 pilot work, we estimated 6 surveys would result in a CV of 7%. Based on the surveys conducted during January 2012, we estimated 5 to 6 surveys would result in a CV of 9% and 10 to 13 would result in a CV of 6%. We show a range in Table A1 because the number of surveys needed to reach a given CV depends on whether data are analyzed using HDS or CDS (see new note in Table A1). We have also updated Table A1 with additional effort estimates. We are recommending that at least 6 surveys be conducted (see Element 3, section Survey repetition and sample size); this does not mean more surveys cannot be conducted. However, our experience suggests that 7 or 8 surveys per year is about the logistical limit that can be accomplished in the 2-week window.

Note, that Butler et al. (2013) recommended being able to detect a –15% annual decline over a 2-year period or a –10% annual decline over a 3-year period to allow enough time to respond to the scenarios. This would require 4% to 6% CV which can be easily obtained by pooling data from multiple years as Reviewer #2 recommended in comment #20.

We think it is worth noting that detection of a 5% annual decline over the course of 2 years would require a CV of approximately 1.5%. Such precision is rather impossible; thus, suggesting that the revised survey detect a 3 to 5% decline annually is also unreasonable. However, detection of a 3–5% annual decline over a 10-year period, instead of a 2-year period, would require a CV of 14–26%. This might be a meaningful alternative to detecting the scenarios outlined in Butler et al. (2013) since this population has grown at an average of 3.9% per year since winter 1938–1939, and has exhibited a 10-year cycle in population growth (Boyce and Miller 1985, Boyce 1986, Butler et al. 2013).

22. Page 6, Line 36-42. Including figure 3. It might be better to show the plot of the raw data as opposed to the selected cut-intervals of 100 m. I assume this was done to improve the model fit but maybe there needs to be some explanation here.

    Frequency data are never shown as raw data but histograms. This Reviewer implies that the model we used to determine detection probability during the winter 2010-2011 surveys was based on binned data but it was not (see Strobel and Butler 2014). Since the draft protocol was reviewed, a Wildlife Society Bulletin paper (Strobel and Butler 2014) has been published that describes the pilot work in great detail. We have added the citation to appropriate places throughout the protocol.

23. Page 6, line 40-42. I think I know what the authors are trying to say here but this seems a case of mixing "apples and oranges". This would be true if the traditional survey technique had fixed transects but because the traditional technique varied transect width and intentionally made excursions off transect to verify bird identification the observer could have counted all the birds and then sum. Because there is no way of quantifying the error associated with the traditional survey

technique it is not correct to say that the detectability was < 1. I think I would recommend deleting lines 40-42 and replace it with a statement like "Because distance from the transect line clearly affects detection then a distance modeling approach is necessary to statistically estimate the number of whooping crane groups that were not detected along the transect."

This comment is incorrect. We show that detectability was less than one (see Strobel and Butler 2014). The assumption that the observers made during the traditional technique was that they could, after multiple surveys, figure out all the individuals in the population (even though most were unmarked). Clearly detectability was less than 100% during an individual survey. Figure 2 shows that for many years the maximum number of whooping cranes counted was often less than the reported abundance; hence, detection less than 100%. For example, the population was reported to have been comprised of 264 birds in 2009, but surveys from that year reported from 201 to 268 birds. Again, there are a lot of misconceptions about the traditional technique and that is why the critique is so important.

24. Page 8, line 1-2. Why would the authors try and gauge distance based on marks on the struts? Even in the traditional census groups were marked on paper DOQQ's.

    The use of marks on struts to measure distance from transect during aerial surveys is a common technique (e.g., Caughley et al. 1976, Guenzel 1997, Butler et al. 2007). That is why we considered it.

25. Page 8, line 21-24. Rewrite to: "Angle of the sun significantly affected detection of whooping crane decoys in our pilot study. Stehn and Taylor (2008) also noted that sun angle affected detection of whooping cranes during the traditions census. To minimize the effect of sun angle we recommend….."

    The sentences in question are "The traditional census was conducted during the morning and afternoon with a break for lunch (Stehn and Taylor 2008). Therefore, much of the survey effort was accomplished when the sun was at lower angles resulting in high detectability on one side of the aircraft but low detectability on the other. We recommend…" We edited those sentences as follows, "The angle of the sun significantly affected detection of whooping crane decoys in our pilot study. Though Stehn and Taylor (2008) also noted that sun angle affected detection of whooping cranes, the traditional census was conducted during the morning and afternoon with a break for lunch. Therefore, much of the survey effort was accomplished when the sun was at lower angles resulting in high detectability on one side of the aircraft but low detectability on the other. We recommend…"

26. Page 9, line 20. Replace "defensible" with "rigorous"

    Change made.

27. Page 11, line 6. Delete the last sentence in the Figure 6 caption. This sentence is not germane to the methods.

    The caption of Figure 6 was "Figure 6. The sampling frame for monitoring whooping crane on their wintering grounds along the Texas gulf coast, USA. These were not formally defined during the traditional "census" effort." We retained this sentence because we want to be clear that a formally

defined sampling frame is new to the revised technique. We believe this sentence is important for the reader to understand how the revised technique compares with the traditional technique. Further, our critique of the traditional technique must be open and transparent in order to increase understanding for why we established the new technique.

28. Page 12, line 2. Delete "scientifically defensible" and replace with "statistically rigorous"

    We deleted "scientifically defensible."

29. Page 12, line 8-12. I understand what the authors are saying here; however, I think reports should include not only those estimates made from the PSF but also legitimate counts from areas well outside of the sampling frame. In addition if qualified observers can verify that they were watching a group of cranes that are outside the sampling frame the entire time a survey was taking place the reports should also add these birds to the peak count.

    We added the following to the new section, "Element 7, comparison with estimates from the traditional technique."

    "Whooping cranes observed outside of the primary sampling frame will be reported separately and identified according to the source of the observation (e.g., Texas Whooper Watch, satellite transmitter, or secondary sampling frame during aerial survey efforts). However, adding these reported birds to the estimate presents potential problems. For example, we cannot ever be completely certain that whooping cranes observed outside the primary sampling frame on one day did not move to or from the primary sampling frame before the surveys were completed."

30. Page 12, line 28. I am not sure what relevance this sentence has to the analyses. In fact, I cannot tell from Appendix C how regions are treated.

    The sentence in question reads, "The sampling units for this design can be considered in a hierarchical or nested sense (Figure 7)." This sentence is simply an introductory sentence to a paragraph that explains how the sampling design was constructed in a hierarchical manner. The regions, primary and secondary sampling frames, are treated differently in that data from the secondary sampling frame are not included in the analysis since so few whooping cranes currently inhabit it. Please refer to the "Element 3, Sampling frame" section among other places throughout the protocol.

    It looks to me like each 1km^2 along the transects are treated as independent samples and all used in the quantification of groups within the PSF.

    Yes, the 1-km^2 grid cells are used as the sampling unit for abundance of groups. However, the groups themselves are the sampling unit for detection probability. Thus, the hierarchical nature of this sampling design.

31. Page 12, lines 32, 38 and 39. The 1-km^2 segments, how will they be characterized? I assume that the entire segment will be given one vegetation class. If this is correct then the 1-km^2 may be too large. I think the TPWD classification effort had a 10-m resolution. (I found later that each 1-km^2 will have 6 classes). This language should be inserted here.

No, each 1-km^2 grid cell will not be given one vegetation class. This is not the appropriate place to discuss the environmental covariates used since this section is about the sampling design not modeling. We inserted a note to refer readers to "Element 4, Environmental covariates" section for more information on the environmental covariates used.

32. Page 15, Box Plot. The authors give good rational for conducting peak abundance surveys during the 4 week period beginning November 28th; however by looking at the box plot it looks like it doesn't matter as long as the surveys are conducted between November 28th and March 27th.

    If we allowed surveys to be conducted over a 4-month period instead of a small temporal window, surveys would be less comparable from year to year. For example, if surveys were conducted during December one year resulting in an estimate of 270 then March the next year resulting in an estimate of 260, one might be tempted to suggest that the population declined between those two years. However, the second estimate accounts for winter mortality (or emigration) whereas the first estimate does not. To ensure comparability from one year to the next, surveys need to be conducted during the same time period. To facilitate comparison with the traditional technique which aimed to provide estimates of "peak" abundance (i.e., abundance post fall-migration prior to any winter mortality), then December is the most appropriate time period.

33. Page 16, line 2-3. I think this sentence is missing something. Maybe it should stay "3-6 surveys will be required to obtain a CV(N) low enough to detect a 10% decline." In addition the stated goal is to be able to detect a 10-15% decline over 3 years that would be a 3-5% decline each year. According to figure A2, regardless of time (2 year or 3 years) the method should be able to detect a 10-15 decline which will require a CV(N) of between .035 and .053.

    See our response to Reviewer's comment #21; we have rewritten the section as per our response in comment #21.

34. Page 17, line 8-21. The statement that the population is approximately 15 years from reaching 400 birds does not seem to be supported by (Butler et al. 2013). By looking a Figure 1 in this citation it looks to me like 400 birds could be attained as soon as 2017 and at the latest 2027. Even the text gives a 50% chance of reaching 400 birds by 2023. It seems likely that we could have 400 birds in the next 7 years; therefore to suggest that monitoring be conducted every other year seems inappropriate. I would say that reaching a potential downlisting criterion in as few as 4 years seems pretty imminent.

    The recovery probability (i.e., probability of reaching 400 birds) did not exceed 80% until 2025. The 15 years out was from winter 2010 (last year of the data used in Butler et al. 2013). We have revised this statement to read "this whooping crane population is ≈12 years from reaching 400 individuals." The Reviewer is correct that we have a 50/50 chance of reaching 400 birds by 2023. However, reaching 400 birds by 2017 is by no means imminent (4% probability).

    I would suggest this section be rewritten to simply state that the monitoring will be conducted annually and that frequency will be revisited with each major revision of this document.

    We choose to keep this discussion about the utility of annual monitoring since it is an important topic one must consider given limited budgets and much priority work. If the precision of abundance estimates from the new technique reach the levels predicted by Appendix A (now that

the bugs have been worked out of the protocol), biennial monitoring must be considered given the reasons discussed in the section.

35. Page 17, line 28. The web link does not work. I suggest you remove the link.

    We have corrected the web link. TPWD recently moved the data. Thank you.

36. Page 17, line 41. Delete the sentence "Tests of marks on struts….." I am not sure why marks on struts was ever explored, even the traditional technique used DOQQ's.

    The use of marks on struts to measure distance from transect during aerial surveys is a common technique (e.g., Caughley et al. 1976, Guenzel 1997, Butler et al. 2007). That is why we considered it. We choose not to delete this sentence.

37. Page 18, line 30. Sentence beginning "However,…." needs a citation, or change the word "will" to "should"

    We have provided citations, thank you.

38. Page 20, line 30. The use of gloves is probably not practical.

    Leather gloves maybe used instead of Nomex gloves; see ALSE Handbook. We changed "Nomex gloves" to "Nomex or leather gloves."

39. Page 21, line 1-7. General Comment: Having a ridged standardized protocol and a good repeatable survey design means that just about anybody should be able to conduct the survey and get similar results. If this is true it might open the door to have the surveys contracted out. I am certainly not advocating this approach now, but down the road once all the kinks are worked out, significant cost savings could be realized.

    Perhaps contracting the survey in the future is possible. However, regardless of the future potential for contracting the work, at least any USFWS employee with a little training will be able to conduct the survey.

40. Page 24, line 24. Change the word "follow" to "following"

    Corrected, thank you.

41. Page 24, line 45. Web link does not work.

    We have corrected the web link. TPWD recently moved the data. Thank you.

42. Page 24, line46-Page 25, line 7. Incorporate this language into Page 17 line 23-33 and Page 12 line 36-42.

    Good idea. We add "The percent of each 1-km^2 transect segment is determined for 6 general vegetation types: saltmarsh, open water, wetland, saltmarsh-shrubland, upland, and urban (Appendix E). Though these covariates are only rudimentary, strong relationships with whooping

crane abundance has been demonstrated with some of them (Strobel et al. 2012)." to the "Element 3, Environmental covariates" section.

43. Page 44, Appendix C. Conduct data analyses and prepare reports. Appendix C Makes no mention of reports and who is responsible for them. Somewhere in this document it should specifically state who will be writing the reports.

    You are correct that preparing reports is not discussed in Appendix C. We have corrected this in the "Overview of the Whooping Crane Abundance Protocol" section. We have also added "Compilation of reports will be a collaborative effort among the Lead Biologist, Regional Biometrician, Regional Data Manager, and Whooping Crane Recovery Coordinator (see Element 8, Roles and responsibilities)." to the "Element 7, Reports" section and the "Element 8, Roles and responsibilities" section.

44. Page 87, Appendix C: General comment: This appendix is well written with the exception of a typo on Page 92, line 2.

    Thank you, the sentence now reads "If a significant (typically $\alpha = 0.15$ is used) negative slope is found, this suggests larger groups tend to be detected at greater distance than smaller ones."

    In the paragraph starting on line 30 of Page 97 the authors recommend pooling data from repeated surveys prior to running the model. Because cranes are territorial groups detected along repeated transects are likely to be correlated. I am not sure how this correlation affects model results but I would think you would want to look at how closely repeated transects are correlated prior to pooling them; however, I am not a statistician.

    We are unsure what the comment is trying to get at. Pooling data from transects is standard practice for distance sampling-based surveys (see Buckland et al. 2001). We pool repeated transects to avoid treating them as if a repeated survey is an independent sample. We do not treat additional surveys of a transect as independent surveys. Instead, we pool data from repeat visits of a transect and treated it as one transect which is the correct way to account for the dependence described in the comment. This ensures we are not inflating the sample size at the transect level.

General Comments:

1. Overall, I am pleased to see the Service move to a standardized method that allows for quantifying error associated with estimating crane abundance.

   Thank you. We are glad to see this Reviewer understands the importance of standardizing survey techniques for whooping cranes.

   I have made numerous suggested changes within this document in regards to the use of the term "scientifically defensible". Although the authors are technically correct to use this language, in this situation, I would refrain from it. In my opinion, the authors spent an inordinate amount of time pointing out flaws in the original census method that were presumably used to justify changing methods. A simple short paragraph detailing that the original census technique lacked a formal repeatable protocol and did not provide quantitative estimates of error would suffice. After all, this is intended to be a written standardized protocol not a justification for changing methods.

We disagree that too much of the protocol was spent on critiquing the traditional technique. As stated before, less than 4 pages of a document containing over 150 pages was devoted to critiquing the traditional technique. Many Reviewer comments revealed misconceptions about the traditional technique which has further emphasized the need for critiquing it. Therefore, we have chosen to include the description and critique of the traditional survey technique. We feel that to "justify changing methods" is an important component of any new survey protocol.

The new protocol is without doubt statistically rigorous, so there is no need to use the term "scientifically defensible". The repeated use of this language throughout this document seems to convey to readers that past practices were not defensible. This is a mistake that should be avoided because the Service is not only judged in courts of law but also in the court of public opinion.

The term "scientifically defensible" was only used twice in the review draft and the term "defensible" was used 6 times. The term "defensible" now only appears twice. The new technique is "scientifically defensible" whereas the old technique was not. It is not a mistake to point out deficiencies in our methods so we can understand them and make corrections. We hope future efforts improve the new techniques when needed.

It is ironic that the very data in question with all its unquantifiable flaws was used by the authors to publish Butler et al 2012 and that publication used in this document to determine precision and suggest less frequent monitoring. This irony will not be lost on other readers.

The traditional technique resulted in 4 to 21 surveys per year (mean = 12.7, SD = 4.2; Butler et al. in prep). We are planning on conducting 6 surveys per year. Sure that is fewer, on average, than the traditional technique but we did not use data from the traditional technique to determine power. We based power decisions on our pilot efforts and the first year of surveys using the new technique. To determine the precision needed, we had to determine the amount of change that needed to be detected. The Butler et al. (2013) paper used data from the traditional technique to better understand this population's dynamics. We are sure the Butler et al. (2013) authors would have preferred to have used data with fewer flaws but nothing else was available. None the less, that work shows a 10 to 15% decline must occur for 2 to 3 years to substantially delay recovery. In the absence of better survey data, we used the best available data to inform our monitoring objectives. Data were used cautiously, based on the assumptions used to acquire them.

2. With any new technique there will be growing pains and, unfortunately, since this method has been used in the last two seasons it has yielded what should be considered unacceptable levels of variation. I think that the CV(N) was around 0.12 in the first year and close to 0.13 in the second. Based on the table provided, on Page 83, it would take a 40 to 50% decline over a 4 year time period before a statistically significant difference could be shown. This certainly does not approach the stated goal of detecting a 10 to 15% decline over 3 years. Much effort should be expended to understand why the technique has performed so poorly. As a person that has been in the plane I can positively state that I have contributed to this variation due to at least one instance of voice recording error. This technique has many moving parts and does require a fair bit of training. I concur with the authors' recommendations for training of new counters and becoming very familiar with the equipment and software prior to conducting surveys. Although training of counters will help decrease variation I do not think it will solve the whole problem.

The first year of the new survey resulted in a CV of 12% and the second year of the new survey resulted in a CV of 19%. Due to unfortunate constraints during the first year, only 3 surveys were conducted that met the protocol criteria. During the second year, new observers misunderstood some of the protocol requirements and data collection procedures which resulted in only one useable survey for that year. Now that all the kinks have been worked out, the full survey effort of 6 surveys should result in the precision discussed in Appendix A (CV $\leq$ 8%) which is enough to detect the change outlined in sampling objective 1.

A more rigorous and in-depth decoy study would help in understanding all of the sources of variation.

The decoy study we conducted was elaborate and much was learned (see Strobel and Butler 2014). We see no reason at this time to conduct another decoy study since all the decoy study does is help us better understand the detection process. Most of the variation in the estimate is due to variation in encounter rates not the detection process. The variation in encounter rates (i.e., the spatial distribution of whooping crane abundance) cannot be effectively studied with a decoy experiment.

3. The technique relies on being able to accurately develop a detection curve, but when there are two observers each may have different detection probabilities. How are these reconciled when half of one transect and therefore, half of each 1 km^2 have different detection probabilities?

   Distance sampling is pooling robust which means that it is extremely robust to variation in detectability due to factors other than distance, such as observers, habitat conditions, and weather (Thomas et al. 2006).

4. Are all the transects of a given observer on the same day used to model the detection curve or are curves modeled for each transect?

   There are not enough detections on a given transect to model a detection curve for each transect nor is modeling transect-specific detection curves normal practice for distance sampling. There are not enough detections on a particular survey day to model a detection curve for each survey either (though we could use survey day as a covariate in the detection model). See discussion about pooling robustness above. We are considering using habitat covariates in the detection models but doubt it will improve the precision of the final abundance estimates much.

5. Were the decoy experiments conducted with two observers? I think much could be learned through an elaborate decoy study.

   Yes, the decoy experiments were conducted with two observers. We added a note in this section that two observers were used. This was an elaborate decoy study and much was learned (see Strobel and Butler 2014).

6. Are each of the 6 surveys within the two week window used as separate samples or is all the data pooled as seems to be suggested in Appendix C?

   All the data are pooled in the analysis which is appropriate since multiple surveys of the same transect cannot be treated as independent samples. See page 79 in Buckland et al. (2001).

7. Finally I would like to commend the authors for developing such and in-depth protocol and series of standard operating procedures. As biologists and ecologists, we should always strive to collect the most meaningful data possible and ensure that our techniques are properly documented and tested.

   Thank you for the complement; this protocol would not have been possible without the contribution of many others, as we now note in the acknowledgements section of the protocol.

Literature Cited:

Bonds, C. J. 2000. Characterization of banded whooping crane winter territories from 1992–93 to 1996–97 using GIS and remote sensing. Thesis. Texas A&M University, College Station, USA.

Boyce, M. S. 1986. Time-series analysis and forecasting of the Aransas/Wood Buffalo whooping crane population. Pages 1–9 in J. C. Lewis, editor. Proceedings 1985 Crane Workshop. Platte River Whooping Crane Maintenance Trust and U.S. Fish and Wildlife Service, Grand Island, Nebraska, USA.

Boyce, M. S., and R. S. Miller. 1985. Ten-year periodicity in whooping crane census. Auk 102:658–660.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, New York, New York, USA.

Burnham, K. P., D. R. Anderson, and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. Wildlife Monographs 72:1–202.

Butler, M. J., W. B. Ballard, M. C. Wallace, S. J. DeMaso, and B. K. McGee. 2007. Aerial surveys for estimating wild turkey abundance in the Texas Rolling Plains. Journal of Wildlife Management 71:1639–1645.

Butler, M. J., G. Harris, and B. N. Strobel. 2013. Influence of whooping crane population dynamics on its recovery and management. Biological Conservation 162:89–99.

Butler, M. J., K. L. Metzger, and G. Harris. Whooping crane demographic responses to winter drought focuses conservation strategies. In preparation.

Canadian Wildlife Service and U.S. Fish and Wildlife Service. 2007. International recovery plan for the whooping crane (*Grus americana*). Third revision. Environment Canada, Ottawa and U.S. Fish and Wildlife Service, Albuquerque, New Mexico, USA.

Caughley, G., R. Sinclair, and D. Scott-Kemmis. 1976. Experiments in aerial survey. Journal of Wildlife Management 40:290–300.

Chandler, R. B., and J. A. Royle. 2013. Spatially explicit models for inference about density in unmarked or partially marked populations. The Annals of Applied Statistics 7:936–954.

Chandler, R. B., J. A. Royle, and D. I. King. 2011. Inference about density and temporary emigration in unmarked populations. Ecology 92:1429–1435.

Conroy, M. J., and J. P. Carroll. 2009. Quantitative conservation of vertebrates. John Wiley & Sons, New York, New York, USA.

Gerrodette, T. 1991. Models of power of detecting trends: a reply to Link and Hatfield. Ecology 72:1889–1892.

Gil-Weir, K. C., W. E. Grant, R. D. Slack, H. H. Wang, and M. Fujiwara. 2012. Demography and population trends of whooping cranes. Journal Field Ornithology 83:1–10.

Guenzel, R. J. 1997. Estimating pronghorn abundance using aerial line transect sampling. Wyoming Game and Fish Department, Cheyenne, USA.

Hedley, S. L., and S. T. Buckland. 2004. Spatial models for line transect sampling. Journal of Agricultural, Biological, and Environmental Statistics 9:181−199.

Hedley, S. L., S. T. Buckland, and D. L. Borchers. 2004. Spatial distance sampling models. Pages 48–70 *in* S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, editors. Advanced distance sampling: estimating abundance of biological populations. Oxford University Press, New York, New York, USA.

Lancia, R. A., W. L. Kendall, K. H. Pollock, and J. D. Nichols. 2005. Estimating the number of animals in wildlife populations. Pages 106–153 *in* C. E. Braun, editor. Techniques for wildlife investigations and management. Sixth edition. The Wildlife Society, Bethesda, Maryland, USA.

Morrison, M. L., W. M. Block, M. D. Strickland, B. A. Collier, and M. J. Peterson. 2008. Wildlife Study Design. Second edition. Sprnger, New York, New York, USA.

Nicol, S., and I. Chadès. 2012. Which states matter? An application of an intelligent discretization method to solve a continuous POMDP in conservation biology. PLoS ONE 7:e28993.

Pugesek, B. H., M. J. Baldwin, and T. Stehn. 2013. The relationship of blue crab abundance to winter mortality of whooping cranes. The Wilson Journal of Ornithology 125:685–661.

Royle, J. A., R. B. Chandler, K. D. Gazenski, and T. A. Graves. 2013. Spatial capture-recapture models for jointly estimating population density and landscape connectivity. Ecology 94:287–294.

Royle, J. A., D. K. Dawson, and S. Bates. 2004. Modeling abundance effects in distance sampling. Ecology 85:1591–1597.

Sollmann, R., B. Gardner, R. B. Chandler, D. B. Shindle, D. P. Onorato, J. A. Royle, and A. F. O'Connell. 2013. Using multiple data sources provides density estimates for endangered Florida panther. Journal of Applied Ecology 50:961–968.

Stehn, T. V. 1992. Unusual movements and behaviors of color-banded whooping cranes during winter. Proceedings of the North American Crane Workshop 6:95–101.

Stehn, T. 2001. Whooping cranes during the 2000–2001 winter. Aransas National Wildlife Refuge, U.S. Fish and Wildlife Service, Austwell, Texas, USA.

Stehn, T. 2004. Whooping cranes during the 2003–2004 winter. Aransas National Wildlife Refuge, U.S. Fish and Wildlife Service, Austwell, Texas, USA.

Stehn, T. V. 2009. Whooping cranes during the 2008–2009 winter. Aransas National Wildlife Refuge, U.S. Fish and Wildlife Service, Austwell, Texas, USA.

Stehn, T. V. 2010. Whooping cranes during the 2009–2010 winter. Aransas National Wildlife Refuge, U.S. Fish and Wildlife Service, Austwell, Texas, USA.

Stehn, T. V. 2011. Whooping cranes during the 2010–2011 winter. Aransas National Wildlife Refuge, U.S. Fish and Wildlife Service, Austwell, Texas, USA.

Stehn, T. V., and E. F. Johnson. 1987. Distribution of winter territories of whooping cranes on the Texas coast. Pages 180–195 *in* J. C. Lewis, editor. Proceedings 1985 Crane Workshop. Platte River Whooping Crane Maintenance Trust and U.S. Fish and Wildlife Service, Grand Island, Nebraska, USA.

Stehn, T. V., and F. Prieto. 2010. Changes in winter whooping crane territories and range 1950–2006. Proceedings of the North American Crane Workshop 11:40–56.

Stehn, T. V., and C. Strobel. 2012. An update on mortality of fledged whooping cranes in the Aransas/Wood Buffalo population. Proceedings of the North American Crane Workshop 12:in press.

Strobel, B. N., and M. J. Butler. 2014. Monitoring whooping crane abundance using aerial surveys: influences on detectability. Wildlife Society Bulletin 38:188–195.

Strobel, B., M. J. Butler, and G. Harris. 2012. Aransas-Wood Buffalo whooping crane abundance survey (2011–2012). U.S. Fish and Wildlife Service, Austwell, Texas, USA. < http://www.fws.gov/uploadedFiles/FY12_ANWR_WWCAS_Review_01OCT12-508.pdf>. Accessed 2 October 2012.

Thomas, L., S. T. Buckland, K. P. Burnham, D. R. Anderson, J. L. Laake, D. L. Borchers, and S. Strindberg. 2006. Distance sampling. Encyclopedia of Environmetrics, John Wiley and Sons. <http://dx.doi.org/10.1002/9780470057339.vad033.pub2>. Accessed 16 January 2014.

U.S. Fish and Wildlife Service. 2013. How to develop survey protocols: a handbook. Version 1.0, National Wildlife Refuge System, U.S. Fish and Wildlife Service, Natural Resource Program Center, Fort Collins, Colorado, USA.

Appendix. R code to demonstrate generalized HDS for a "full-coverage design."

```
library("MASS")
library("unmarked")

R=10 # number of transects
tr=6  # number of replicates
sw=50 # Strip width (transect half-widt).
tl=sw*2*R # Transect length.
Q=tl/(sw*2)
# This results in 10 transects that are 100 m wide over a landscape that
# is 1,000 x 1,000 m.  Therefore, this landscape would be composed of
# 100 grid cells that are 10,000 m².

# Average abundance/grid cell for the two habitat types.
lambda1=2.0
lambda2=0.5

sigma=30 # Half-normal shape parameter for the detection curve.

# Home range shape.
hr.shape=matrix(c(500,0,0,500),2,2)

# So we want animals to be moving in and out of each cell. This plot
# demonstrates the posible distances moved between surveys.
plot(mvrnorm(1000,c(0,0),hr.shape))

# Function to locate each animal during a survey.
location=function(n,center){mvrnorm(n,center,hr.shape)}

it=1000 # Number iterations.

# Make a data frame to hold results.
A=as.data.frame(matrix(NA,it,6))
names(A)=c("phi","lambda1","lambda2","abun.pred","abun.est","abun.true")

# Start the simulation loop.
for (B in 1:it){

  # Make grid cell centers.
  cell.centers=cbind(sort(rep(seq(sw,R*sw*2-sw,sw*2),R)),rep(seq(sw,tl-
      sw,sw*2),R))

  # Split the landscape into 2 habitat types.

      cell.centers=cbind(cell.centers,sample(c(rep(1,floor((R*Q)/2)),rep(0,ce
      iling((R*Q)/2)))))

  # Make an empty column to add lambdas.
  cell.centers=cbind(cell.centers,matrix(NA,R*Q,1))

  # Assign lambdas to their habitat types.
  for (i in 1:(R*Q)){
    if (cell.centers[i,3]==1)
      cell.centers[i,4]=lambda1
    else
      cell.centers[i,4]=lambda2
  }

  # Make cell.centers a data frame and name variables.
  cell.centers=as.data.frame(cell.centers)
  # Add an empty column to add the randomly selected abundance for each cell.
  cell.centers=cbind(cell.centers,matrix(NA,R*Q,1))
  names(cell.centers)=c("c.x","c.y","type","lambda","n")
```

```
# Randomly select abundance for each cell based on Poisson distribution.
for (i in 1:(R*Q)){
  cell.centers$n[i]=rpois(1,cell.centers$lambda[i])
}

# Make a data frame to hold the home range centers for each animal in the
    population.
home.center=as.data.frame(matrix(NA,1,5))
names(home.center)=c("c.x","c.y","type","lambda","n")

# Fill up the home range centers data frame.
for (i in 1:(R*Q)){
  if (cell.centers$n[i]>0)
    for (j in 1:cell.centers$n[i]){
      home.center=rbind(home.center,cell.centers[i,1:5])
    }
  else
    home.center=rbind(home.center,cell.centers[i,1:5])
}
home.center=home.center[-1,]

# Make two empty columns to hold the x and y for the home range centers.
home.center=cbind(home.center,matrix(NA,length(home.center$n),2))
names(home.center)=c("c.x","c.y","type","lambda","n","hr.x","hr.y")

# Randomly select the home range centers.
for (i in 1:length(home.center$n)){
  if (home.center$n[i]>0)
    home.center$hr.x[i]=home.center$c.x[i]+runif(1,-45,45)
  else
    home.center$hr.x[i]=NA
  if (home.center$n[i]>0)
    home.center$hr.y[i]=home.center$c.y[i]+runif(1,-45,45)
  else
    home.center$hr.y[i]=NA
}

# Make a plot to visualize where the animals in the population live.
#
    plot(home.center$c.x[home.center$type==1],home.center$c.y[home.center$t
    ype==1],pch=18,col="red",xlim=c(-50,R*sw*2+sw),ylim=c(-50,tl+sw))
#
    points(home.center$c.x[home.center$type==0],home.center$c.y[home.center
    $type==0],pch=18,col="blue")
#
    points(home.center$hr.x[home.center$type==1],home.center$hr.y[home.cent
    er$type==1],pch=4,col="red")
#
    points(home.center$hr.x[home.center$type==0],home.center$hr.y[home.cent
    er$type==0],pch=4,col="blue")

# A data frame to hold animal locations during each survey.
movement=cbind(home.center,matrix(NA,length(home.center$n),tr*2))

# Just some code to name the columns.
x=factor(seq(1,tr,1))
aa=NA
for (i in 1:length(x)){
  aa=c(aa,paste("x.",x[i],sep=""),paste("y.",x[i],sep=""))
}
aa=aa[-1]
names(movement)=c("c.x","c.y","type","lambda","n","hr.x","hr.y",aa)
```

```
# Randomly select each animals location during a survey; location based on
    home range center and shape.
for (i in 1:length(movement$n)){
  for (t in 0:(tr-1)){

    movement[i,(t+t+8):(t+t+9)]=location(1,c(movement$hr.x[i],movement$hr.y
    [i]))
  }
}

# Make a plot to visualize where the animals in the population live and
    their movement.
#
    plot(home.center$c.x[home.center$type==1],home.center$c.y[home.center$t
    ype==1],pch=18,col="red",xlim=c(-50,R*sw*2+sw),ylim=c(-50,tl+sw))
#
    points(home.center$c.x[home.center$type==0],home.center$c.y[home.center
    $type==0],pch=18,col="blue")
#
    points(home.center$hr.x[home.center$type==1],home.center$hr.y[home.cent
    er$type==1],pch=4,col="red")
#
    points(home.center$hr.x[home.center$type==0],home.center$hr.y[home.cent
    er$type==0],pch=4,col="blue")
# for (i in 0:(tr-1)){
#   points(movement[,(i+i+8)],movement[,(i+i+9)],pch=20)
# }

# Remove all locations that are outside of the sampling frame.

    movement[cbind(matrix(TRUE,length(movement$type),7),(movement[,8:(7+tr*
    2)]>0))==FALSE]=NA

    movement[cbind(matrix(TRUE,length(movement$type),7),(movement[,8:(7+tr*
    2)]<(sw*2*R)))==FALSE]=NA
for (i in 0:(tr-1)){
  movement[is.na(movement[,(i+i+8)]),(i+i+9)]=NA
  movement[is.na(movement[,(i+i+9)]),(i+i+8)]=NA
}

# Make a plot to visualize where the animals in the population live and
    their movement.
#
    plot(home.center$c.x[home.center$type==1],home.center$c.y[home.center$t
    ype==1],pch=18,col="red",xlim=c(-50,R*sw*2+sw),ylim=c(-50,tl+sw))
#
    points(home.center$c.x[home.center$type==0],home.center$c.y[home.center
    $type==0],pch=18,col="blue")
#
    points(home.center$hr.x[home.center$type==1],home.center$hr.y[home.cent
    er$type==1],pch=4,col="red")
#
    points(home.center$hr.x[home.center$type==0],home.center$hr.y[home.cent
    er$type==0],pch=4,col="blue")
# for (i in 0:(tr-1)){
#  points(movement[,(i+i+8)],movement[,(i+i+9)],pch=20)
# }

# Make a data frame to hold distance data.
locs=as.data.frame(matrix(NA,1,5))

# Fill that data frame.
for (i in 0:(tr-1)){
```

```
  x=(findInterval(movement[,(i+i+8)],seq(0,tl,sw*2))-
    findInterval(movement$c.x,seq(0,tl,sw*2)))*100
  y=(findInterval(movement[,(i+i+9)],seq(0,tl,sw*2))-
    findInterval(movement$c.y,seq(0,tl,sw*2)))*100
  x[is.na(x)]=0
  y[is.na(y)]=0

    locs=rbind(locs,(cbind(movement$c.x+x,movement$c.y+y,i+1,movement[,(i+i
    +8)],movement[,(i+i+9)])))
}
names(locs)=c("c.x","c.y","time","x","y")
locs=locs[-1,]

# Assign habitat types.
for (j in 1:length(locs$c.x)){

    locs$type[j]=cell.centers$type[(cell.centers$c.x==locs$c.x[j])&(cell.ce
    nters$c.y==locs$c.y[j])]
}

# Calulate distance from transects.
locs$dist=abs(locs$c.x-locs$x)

# Estimate detection probability.
locs$p=exp(-locs$dist^2/(2*sigma^2))

# Determine if it was detected or not.
locs$det=rbinom(length(locs$c.x),1,locs$p)
locs$det[locs$det==0]=NA
locs$det.dist=locs$det*locs$dist
# Make cell IDs.
locs$ID=paste(locs$c.x,locs$c.y,sep="")

# Get a list of unique cell IDs.
ids=unique(locs$ID)

# Distance breaks.
breaks=seq(0,50,by=10)

# An array to put the data in.
y=array(NA,c(R*R,length(breaks)-1,tr))

# Fill the array.
for (t in 1:tr){
  for (i in 1:length(ids)){

    y[i,,t]=table(cut(locs$det.dist[(locs$time==t)&(locs$ID==ids[i])],break
    s,include.lowest=TRUE))
  }
}

# Convert array to matrix
y=matrix(y,nrow=R*R) # convert array to matrix

# Organize the data for analysis with unmarked.

    umf=unmarkedFrameGDS(y=y,survey="line",unitsIn="m",dist.breaks=breaks,t
    length=rep(sw*2,R*R),numPrimary=tr,siteCovs=data.frame(type=cell.center
    s$type))

# Fit the generalized HDS model.
m1=gdistsamp(~type,~1,~1,umf,output="density")

# Get lambda estimates from the model.
```

```
  A[B,2:3]=exp(coef(m1,type="lambda"))

  # Get phi estimate from the model.
  A$phi[B]=(exp(coef(m1,type="phi"))/(1+exp(coef(m1,type="phi"))))

  # Get predicted abundance (unadjusted) from the model.
  A$abun.pred[B]=sum(predict(m1,type="lambda")$Predicted)

  # Get predicted abundance (adjusted) from the model.
  A$abun.est[B]=sum((predict(m1,type="lambda")$Predicted)*A$phi[B])

  # Get true abundance of the population.
  A$abun.true[B]=sum(cell.centers$n)

}

# True abundance.
true.abundance=(lambda1*(R*Q)/2)+(lambda2*(R*Q)/2)

# Estimate Bias.
A$Bias=(true.abundance-A$abun.est)/A$abun.true
mean(A$Bias)
quantile(A$Bias,probs=c(0.025,0.975))

# Look at mean lambdas.
mean(A$lambda1)
quantile(A$lambda1,probs=c(0.025,0.975))
mean(A$lambda2)
quantile(A$lambda2,probs=c(0.025,0.975))

# Look at the adjusted abundance estimates.
mean(A$abun.est)
quantile(A$abun.est,probs=c(0.025,0.975))

# Look at the adjusted abundance estimates.
mean(A$abun.pred)
quantile(A$abun.pred,probs=c(0.025,0.975))
```