A PEER REVIEW OF THE ARCTIC PRISM PROGRAM

Recommended Citation: Arctic PRISM Peer Review Committee. 2010. A Peer Review of the Arctic PRISM Program. U.S. Shorebird Conservation Plan, U.S. Fish and Wildlife Service, Division of Migratory Bird Management, Arlington, VA, USA. Available online at <http://www.fws.gov/shorebirdplan/Prism.htm>.

Table of Contents

INTRODUCTION

During a meeting in Quebec City, Quebec (15-16 October 2002), shorebird biologists from Canada and Alaska met to discuss proposed methodologies for conducting breeding shorebird surveys in the Arctic. Jon Bart provided background information on methods that he and Brad Andres began developing in Alaska in 1998. Richard Lanctot described potential problems with the protocols that other shorebird biologists had relayed to him, and suggested it would be beneficial to have the methodology peer-reviewed prior to implementation. These issues were reviewed by a subcommittee (Richard Lanctot (chair), Jon Bart, Brad Andres, Stephen Brown and Guy Morrison), and the information was presented at the Waterbirds Meeting in LaCrosse, Wisconsin, 4-6 November, 2002. At the meeting, Jon Bart provided additional information on the Arctic surveys and presented a proposal for completing them. At this time, it was formally agreed that a peer review be conducted and Bruce Peterjohn volunteered to serve as chairperson. Bruce prepared a document (Appendix 1) outlining the peer review process. However, before methods could be peer reviewed, it was necessary to prepare a document providing details of the statistical background, methodology and proposed implementation of Arctic PRISM. This document was completed by November, 2004.

By December, 2004, the document and a series of 27 questions was provided to six external reviewers that specialized in statistics and/or shorebird breeding ecology. These questions were generated primarily by Richard Lanctot, Bruce Peterjohn and Jon Bart. Reviewers could also provide general comments as they wished at the end of the document (question 28). Comments were received from each reviewer between February and April, 2005. At a meeting in Galveston, Texas, February, 2005, Stephen Brown, Vicky Johnston, Jon Bart, Paul Smith and Rick Lanctot divided the questions up, and each person collated and summarized comments from all reviewers for their particular set of questions (see "summary of reviewers' comments" below each of the 28 questions in the following text). Responses to these comments were prepared by Stephen Brown, Vicky Johnston, Jon Bart and Paul Smith (see "response to reviewers"), but questions of a statistical nature were addressed primarily by Jon Bart.

Over the next two years, the committee wrestled with how best to handle responses to reviewers. By February 2007, Bruce Peterjohn had identified a number of key issues he felt needed to be addressed before the peer review would be completed. The process was resurrected in September 2009, when Jon Bart and Paul Smith addressed the last remaining issues identified by Bruce Peterjohn in his final assessment.

The delay was in large part due to a continual evolution of the PRISM methods. In the years since the external review was initiated, Arctic shorebird surveys had continued in Alaska and Canada. Consequently additional data had been collected in many areas where no information was available before. In fact, so much additional information was collected that Jon Bart was able to revise the original analyses. The original power analysis relied on many subjective decisions about densities and expected population sizes because of sparse data. Because survey coverage was now much more extensive, it was possible to use neighboring regions as surrogates for unsampled regions. This greatly simplified the analyses, and addressed a number of substantive comments from reviewers. Other issues raised by reviewers were addressed through revisions of the original document provided to the reviewers, which has recently been finalized and submitted to Studies in Avian Biology.

In the document below, each question that was provided to the peer reviewers is presented in sequence, followed by their responses, and a summary of their responses. Finally, a response to the reviewers is provided at the end of each question. In each case, we indicate if a question is still relevant to the new analysis and methods, and if so, our responses to reviewers' comments.

The original questions and summaries of reviewers' comments were prepared by the Arctic PRISM Peer Review Committee, which consisted of Bruce Peterjohn (chair), Brad Andres, Jon Bart, Stephen Brown, Richard Lanctot, Vicky Johnston, and Paul Smith. The "response to the reviewers" was generated primarily by Jon Bart, Stephen Brown, Vicky Johnston and Paul Smith.

QUESTIONS AND RESPONSES

1. MONITORING STANDARD

Is the monitoring standard of "80% power to detect a 50% decline during no more than 20 years using a two-tailed test, a significance level of 0.15, and acknowledging effects of potential bias" reasonable given shorebirds are long-lived and these population changes would occur within 1-2 generations for many species? If not, can you suggest a more appropriate standard? (Note: if you choose to suggest a different standard please review Bart et al. JWM 68:611-626 and explain why you disagree with the rationale they present.)

Q1: Reviewer #1

The use of a significance level of 0.15 implies the authors are willing to accept a 15% false positive rate i.e. if there were 20 species exhibiting no population change then they are likely to declare falsely that 3 of them are significant. Presumably declaring a significant difference would then direct future research toward finding the cause and management efforts toward reverse the decline. When these efforts are caused by a false positive they are a waste of resources which would be better put to use towards species which are actually declining but this may be an acceptable cost introduced because of a desire to avoid false negatives.

Setting the power at 80% implies the authors are willing to accept a 20% chance of a false negative. Thus even if there has been a 50% decline in the population there is a 20% change the statistical test will fail to declare the result significant. This is somewhat at odds with the significance level being set unusually high (0.15). Increasing the statistical significance to 0.15 is a indication that failing to detect a trend has serious consequences while setting the power at only 80% downplays these consequences. The authors appear to have chosen the increase both the type I and II error rate above the standard levels because they had a premonition to what sample size could be attained rather than setting criteria by assessing the consequences in terms of error rates.

The concept of a 50% decline in 20 years doesn't reflect the actual design being proposed. The design seems to be based on running the survey in two separate time periods and then testing whether there has been a 20% decline between the two periods. The two survey periods are not necessarily 20 years apart. In fact one scenario has the second period starting immediately after for first finishes. A more appropriate phase would be a 50% decline between the two survey periods.

I don't see why the phrase "acknowledging effects of potential bias" is included in the sentence. There is no discussion of potential bias and no bias term is included in the equations. Is the estimation of detection rate how potential bias is acknowledged?

In general an estimate with a CV of 0.30 seems too large to be of any scientific value. For example a population of 100K would have a crude 95% confidence bounds (±2SD) of 40K -160K. I would feel that CV of 0.1 to 0.15 would be more reasonable.

Q1: Reviewer #2

I tend to feel that these standards are reasonable. When I first heard of them I had some of the same concerns that I know others have raised, but having heard Jon explain the rationale in detail, I am persuaded that they have been thought out in great detail and with much attention paid to both the biology and to what can be realistically achieved. The standards are certainly not perfect, and are open to criticism, but any such standard has an arbitrary component and so arguments could be leveled at whatever is chosen. Moreover, since the standard is somewhat arbitrary the things that determine what is "reasonable" are largely not scientific – rather the key decisions are: (1) what level of decline are we prepared to miss, and (2) what resources are we prepared to spend.

The major concern people will probably have is the use of $\alpha = 0.15$ rather than $\alpha = 0.05$. This concern arises, I believe, as a result of history and convention rather than any objective scientific evaluation of what is an appropriate α-level. As many have pointed out before, there is nothing magical about the common use of $\alpha = 0.05$ in ecology. Indeed (I am told) that in other scientific fields much higher burdens of proof are required for something to be viewed as "significant". Equally, as a statistician once pointed out to me, in our court system we are sometimes willing to go with a much lower burden of proof (e.g., in civil cases a simple majority …. akin, in a way, to $\alpha = 0.50$ …. is enough). Of course, it would be great to reduce the risk of Type I errors by having a smaller α, but without a concomitant increase in sample size, that would increase the risk of Type II errors. And I tend to side with the idea that when you are monitoring something, it is better to find an effect when there isn't one, than to conclude no effect when one exists. These standards seem to provide a reasonable balance within the constraints of what sampling is likely to be achievable.

Q1: Reviewer #6

Yes. This scheme focuses on biologically meaningful results. The t-test provides a statistical measure of the significance of a biologically meaningful change. Your words "50% decline" exclude any interest in knowing about a 50% increase in a species. I would think that if a species happens to increase you would want to know about it!

Q1: SUMMARY OF REVIEWERS' COMMENTS

Of the six reviewers, four felt that the overall monitoring standard was reasonable, two (#1 and #2) thought power should be higher, and one thought that demographic rates, rather than density, should be monitored. Reviewer #1 also noted that the significance level is unusually high, though he did not object to this level. All four of the reviewers who supported the overall monitoring standard agreed that, while a higher standard might be desirable, the tradeoff between higher standards and available resources was appropriate.

Q1: RESPONSE TO REVIEWERS

We have two responses to the comment made by Reviewer #1 that the power target should be higher (e.g. 90% power to detect a 50% decline). First, and mainly, 80% is the target adopted by the shorebird community after extensive discussion (Skagen et al. 2003). It would be inappropriate for us to choose a different target. Second, raising the target power or decreasing the target decline would greatly increase the cost of any monitoring program regardless of whether it used the Arctic PRISM methods. We doubt that managers would be willing to spend the needed funds. Our response to the comment by Reviewer #2, that we should try to detect a smaller decline is similar: we address the target adopted by the shorebird community. We note however, that the plan proposed is designed to achieve the accuracy target for some rarer species. The target will be surpassed for more common species, allowing for identification of smaller declines. Our power analysis suggests that a $CV \leq 0.20$ will be achieved for 10 species. That said, we have never seen significant resources expended for a non-game species until the decline had substantially exceeded 50%. That is part of the rationale for choosing this target.

Skagen, S., J. Bart, B. Andres, S. Brown, G. Donaldson, B. Harrington, V. Johnston, S.L. Jones, and R.I.G. Morrison. 2003. Monitoring the shorebirds of North America: towards a unified approach. Wader Study Group Bulletin 100:102-104

Reviewer #1 was also concerned that the use of a significance level of 0.15 would result in a higher than desirable rate of false positive detections of significant declines. The risk identified was the possible misdirection of resources toward causes of declines that were not actually occurring. Our experience, however, has been that when a serious decline is reported, the first action taken in response is more study to verify or refute the initial suggestion. In addition, the general philosophy is to be conservative, and assume that the effects of false negatives would be worse than false positives when detecting declining trends. In fact, reviewer #5 specifically elaborated this philosophy. For these reasons, the overall balance was established to provide a higher likelihood of providing information about significant negative declines, with corroborating results from other types of surveys as a primary safeguard against investing too heavily in determining the causes of declines that are not actually occurring.

We do not advocate that arctic PRISM be the sole source of population information for shorebird monitoring in North America. While we believe that arctic PRISM will perform a crucial role in a continent wide monitoring program, several different surveys across species' ranges, completed independently, would provide a much more solid foundation for management decisions. We hope that arctic PRISM is supporting the development of a broader monitoring program by accepting this modest target for power and recognizing that these surveys should be complimented by others across the range.

Reviewer #2 argued that an entirely different approach should be taken to determine population trends for shorebirds, specifically monitoring of demographic parameters. The reviewer suggests that a 3.4% annual rate of decline could be detected in 3-5 years by monitoring adult survival rates. We recognize that this approach has several strengths, but feel that it would be best employed as a complementary survey method for target species in specific locations of concern. The arctic PRISM surveys proposed here would help to identify the species and appropriate locations to do such studies.

One of the major strengths of the arctic PRISM program is that it will generate population information from low density areas and marginal habitats; habitats which may in fact be the first to experience declines. We feel that it would be impractical to monitor demographic rates for numerous species across these vast, low density areas. For even the most common shorebird species, important parameters such as rates of fledging and juvenile survival have proven difficult to measure. A primary reason for employing demographic monitoring is to identify the life history stage responsible for population declines, and substituting literature values for these difficult to measure parameters negates to some extent the benefits of demographic monitoring.

We were asked recently to elaborate on our view that obtaining reliable estimates of population trend by monitoring demographic rates, at least for shorebirds in the Arctic, is impractical. The issue is explored in detail, including estimates of the needed sample sizes, in:

Bart, J., S. Brown, R.I.G. Morrison, and P.A. Smith. (Submitted) Other Methods for Estimating Trends of Arctic Birds. Appendix A in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

We also note that question 27 asked specifically if a demographic approach should be considered as an alternative and all of the reviewers but reviewer #2 felt that measuring vital rates would be more difficult and less effective than monitoring population size (although several suggested that the additional information gained on reasons for declines would be nice to obtain).

We do agree, however, that in some cases demographic information can be valuable both for identifying declines and for determining which life history stages are affected. In recent years, we have designed the intensive plot work so that estimates of hatch success can be obtained. We also advocate the creation of long term monitoring sites, Tier II sites, where demographic information would be collected at fixed intervals. We invite proposals for how demographic monitoring might be employed to monitor species which are identified in the power analysis as difficult to monitor with our proposed sampling scheme. It is our opinion however that the method we propose is a more cost effective solution to monitoring a large number of species across a broad geographical area.

2. CV THRESHOLD

Is the threshold $CV\left(\hat{Y}\right) \leq 0.31$ *correct? Is the development of this target adequately supported by the expressions provided in the Appendix? If not, can you suggest a more appropriate approach for developing this threshold?*

Q2: Reviewer #1

The development of the threshold follows logically. However there is an assumption that the sample size is large enough so that the normal theory is appropriate. No effort at assessment of whether this is valid has been presented. The design provides $CV(\hat{Y}) < 0.31$. This coupled with the fact that $\hat{Y} \geq 0$ implies that the estimate will be bounded far enough away from zero that a symmetric distribution is at least plausible. A simulation studies might be reasonable to check the validity of the assumption.

The adjustment to the accuracy target when only a portion of the species is breeds in the arctic hasn't been explained well. The result is reasonable but need some further explanation. Possibly the authors should add an appendix with the following argument. Let $\hat{Y}_a$ denote the estimated population in the arctic and $\hat{Y}_b$ denote the estimated remainder of the population then

$$CV(\hat{Y}_a + \hat{Y}_b) = \sqrt{CV(\hat{Y}_a)^2 f_a^2 + CV(\hat{Y}_b) f_b^2}$$

where $f_a = \dfrac{Y_a}{Y_a + Y_b}$ is the estimated proportion of the population in the arctic

Setting a design to $CV(\hat{Y}_a)/\sqrt{f_a}$ will result in an overall estimate with the appropriate CV assuming that a comparable effort is made in the area outside the arctic.

Q2: Reviewer #2

Cursory inspection suggests that calculations are correct, but I leave this question for statisticians to evaluate.

Q2: Reviewer #3

Given that the estimates of mean and SE are unbiased and do not change with population size, the derivation in Appendix 1 seems correct. A particular CV target is valuable in that it provides a measuring tool to compare sampling designs and allocation options. Over time, better data on habitat, effects of annual weather conditions, age structure, or interactions with other species will identify variables for the regression models that may reduce the SEs. Typically variance is less well estimated than the mean.

Q2: Reviewer #4

I had difficulty following the appendix. It seems to me that for a 15% 2-tailed test, one wants (as stated in the appendix) $|r-1|/se(r) = Z_{0.925} = 1.44$, where $r = log(y_2)-log(y_1)$. With r = 0.5, se(r) = 0.5 / 1.44 = 0.35. The $var(r) = [se(r)]^2 = (1)^2 var[log(y_2)] + (-1)^2 var[log(y_1)] = 2 var[log(y)]$. Then se(r) = 0.35 = 1.41 se[log(y)] or se[log(y)] = 0.25.

If one measures the change as a difference instead of a ratio, se(y) = 0.25. If r=0.5, then cv(y) = 0.50.

This approach and the one in the plan do not consider the use of double sampling, which will introduce another source of variability. Consequently, the required sample size may be substantially underestimated.

Q2: Reviewer #5

I've left this question to the statisticians.

Q2: Reviewer #6

I cannot really address whether this is correct or not – only time will tell. Your development of this in the Appendix is correct and your assumptions are warranted. Your analysis certainly indicates that this expression is probably sufficiently close for all practical purposes. Considering the large suite of species that you are targeting, the choices and assumptions that you have made are necessary to support the compromises that you had to make.

Q2: SUMMARY OF REVIEWERS' COMMENTS

Four reviewers noted that the development of the CV target was appropriate, and two did not comment directly on this question but raised specific concerns. Reviewer #1 questioned whether the sample size is large enough to justify the use of normal theory. Reviewer #4 had difficulty with the appendix and suggested that the use of double sampling would require a substantially larger sample size.

Q2: RESPONSE TO REVIEWERS

Reviewer #1 pointed out a mistake in the formula for estimated population size. It implies that separate detection rates are being estimated for each stratum whereas in fact they are estimated by combining results across strata. This greatly increases the sample size and meets guidelines (in Cochran 1977) for when the large sample theory may be used. We appreciate the reviewer pointing out this error and have corrected the formula. Reviewer #4 defines r as $\log(y_2)$- $\log(y_1)$ but we define r as y_2/y_1 (see Bart and Smith, submitted). We believe this is why he arrived at a different CV. He also questioned whether we acknowledged uncertainty about the true detection rates in calculating the variance of the estimated population size. The answer is that we do, for example in expressions (8) and (10).

Since the peer review was initiated, we have explored the threshold CV in more detail. In the original manuscript, we assumed that the CV of the second population estimate was equivalent to that of the first, and that new plots were revisited in the second round of surveys. In the revised analyses, we have investigated the gains in power that could be achieved by revisiting plots, and have explored how the CV of the second population estimate might vary as a function of CV(Y_1) and rate of decline. For additional details, see:

Bart, J., and P.A. Smith. (Submitted) Design of future surveys. Chapter 13 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

3. SURVEY INTERVAL

The manuscript recommends conducting each survey during either "a relatively short period, such as 4-6 years" or "about a decade with the second survey period beginning immediately after the first period ends." Recognizing the availability of funds may dictate the length of survey periods and intervals between surveys, are both approaches appropriate for making valid inferences about population change?

Q3: Reviewer #1

The survey is designed to detect a 50% decline between the two survey periods while the estimates in each period are averages over several years. It is necessary that the midpoints of the two time periods be separated by at least a decade and that each period be constrained to as few as possible (preferably 4 years). If the project is run with the second time period starting once the first has been completed then some more effective analysis which can incorporate strata and year effects should be considered. Other wise it becomes implausible that a 50% decline would occur.

Q3: Reviewer #2

The power to detect a population trend will depend upon the frequency with which population estimates are generated. This has not been taken into account in these documents. The power analysis in the manuscript is predicated on comparing two independent estimates of population size. The goal is to be able to detect a 50% change in population size after 20 years have elapsed. If an estimate takes 5 years to generate, then the second population estimate for 20 years elapsed time will not be available for comparison until the end of the second 5-year survey period, which will really be 25 years after the program has first started. It is at this point that the power analyses will apply. If the second round of surveys is begun sooner (e.g., either in year 6 or in year 10), this will add to the power to detect a change in population size, but one could not expect to detect the targeted rate of change at the end of those surveys. One could calculate the estimated increase in power to detect a certain level of population change with increasing frequency of the repeated surveys, but that has not been presented in these documents.

If each round of surveys takes 10 years to complete, then the power to detect the targeted amount of change will not be achieved until 30 years after the first round of surveys was initiated. The longer it takes to conduct a round of surveys to generate a population estimate, the longer it will be before an estimate of population change will be available that meets the specified monitoring goal.

Replicating survey plots instead of surveying a new, independent sample of plots would increase greatly the power to detect trends. The power will also be highly dependent on the frequency with which the plots are resurveyed. One would need some measure of inter-annual variation in numbers of birds across a sample of plots to determine how much the power to detect trends would be increased.

Q3: Reviewer #3

Yes. Any period, or time interval between groups of years, is valid for an estimate of the average annual rate of change.

I am more concerned that the period over which the data are collected is long enough to average over the sometimes extreme differences between years in arctic environments. As an oversimplified example, care should be taken to avoid the chance that data in 3 good nesting years are being compared with data from 1 good and 2 poor years. Deviation from assumed constant territory occupancy (breeding effort) may amplify or mask changes in actual population size if weather conditions in some years or regions essentially exclude breeders from either coming to or staying on breeding habitat that may be suitable in other years. Regional weather and other covariate data should be actively investigated to try to understand and perhaps reduce such potential biases.

The chance for funding, opportunistic sampling, and methods to staff or administer the proposed PRISM program are simply not discussed and cannot be considered here.

Q3: Reviewer #4

Both approaches are appropriate, given that availability of funds may dictate the length of survey periods and intervals between surveys. It may be helpful to revisit the same sites to obtain the variance reduction of a paired t-test, which removes the site component of variance.

One will want to know if there is a precipitous decline before the end of the second period. At any time, one can conduct an unpaired t-test for any two periods.

Q3: Reviewer #5

I don't feel I have adequate information (or perhaps personal knowledge) to address this question. My concern is that much of any decline will be happening within the period over which each population estimate is being made, and thus not be detected as part of the estimated decline. This is because the trend is estimated from the difference between the abundance estimates for the two periods, yet those abundances are effectively/approximately the "average" abundance across the years of the survey. The t-test will only address the difference between these averages, not that between the start of the first survey period and the end of the second; which is what you would really want to have.) The longer the period for each survey, the greater this problem would be. Therefore, one would like to know what portion of the decline will go undetected for different combinations of survey period and population trajectory, and then see how these things trade-off. It seems like this shouldn't be hard to determine via simulation (there's probably an analytical solution too but you'd have to ask a better mathematician than I).

Q3: Reviewer #6

Both approaches are valid for making inferences about population change, however undetectable changes may exist between repeat surveys that would be detectable if surveys were spaced further apart in time. Preferable would be a relatively short survey period (1 year would be ideal!!) and repeated after some sort of gap, e.g., 5 years. This would depend on what trend you would expect to encounter that is suggested from other sources.

Q3: SUMMARY OF REVIEWERS' COMMENTS

The reviewers generally felt that both approaches were valid, but one reviewer suggested that if the second round of surveys begins as soon as the first is over, then a different analysis should be used. The reviewers also pointed out that if completing two rounds of surveys takes longer than 20 years, then the power objective may not be achieved until the end of round two.

Q3: RESPONSE TO REVIEWERS

We agree with the criticism that under the current design less than 20 years will elapse between consecutive surveys in a region whereas we assumed, in the analysis, that the interval would be 20 years. Another problem with the previous power analysis was that it assumed a new sample of plots would be surveyed in round two, whereas re-surveying the same plots would substantially improve precision. We addressed both issues in a completely new analysis of power.

We also note that regional trend information would be available sooner; as soon as regions are revisited for surveys. Power to detect trends may be low for these regional estimates, however, because the program was designed to meet the accuracy target for the range wide surveys.

Although an excerpt appears below, the issue of rotation time and power to detect trends is explored in detail in:

Bart, J., and P.A. Smith. (Submitted) Design of future surveys. Chapter 13 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

> *Surveys in Alaska could be conducted in a few years so that the 20-year spacing is probably realistic. On the other hand, it seems likely that managers will not want to wait 20 years for a second survey. They will probably conduct the survey roughly once a decade. This means that within ~20 years, three surveys will have been conducted. At present, it is difficult to predict how much value will be added by the second survey or how an analysis using all three surveys will be conducted (regression with 3 data points is usually not recommended). Thus, the second survey will certainly add considerable value, but no reliable way exists to predict how much it will increase precision. The situation in Canada is more complex because a round will take 5 - 10 years. It seems likely that if the second round shows declines, then a third round will be undertaken more rapidly to determine if the apparent declines have continued.*
>
> *In the face of this complexity, it seems that two results are useful in deciding how large a sample of surveys is needed. One is the number of species for which the accuracy targets are achieved for different total sample sizes. The other is the CV for each species and sampling plan. We present both of these statistics below…The results showed that as the number of crew-years increased from 30 to 50 the number of adequately covered species (out of 26) increased from 19 to 25.*

4. DEFINITION OF NUMBER OF BIRDS ON A PLOT

Determining the number of birds on a plot (defined as "the number of territorial males, and their mates, whose first nest of the season, or territory centroid for non-nesters, is within the plot") may be difficult for some species. Examples include (1) non-monogamous species (e.g. a red phalarope female may represent between 1-5 nests); (2) species with segregated feeding and nesting territories; (3) species that re-nest; (4) species with overlapping nesting territories; (5) distinguishing territorial individuals from non-breeders; (6) secretive species; and (7) those with very large territories. How do these uncertainties influence the ability of PRISM to detect population change under the proposed sampling plan? Does this seem like a suitable parameter for surveyors to estimate? If not, can you identify another one? Is it valid to assume that the PRISM estimate of population size provides an unbiased estimate of the true number of birds?

Q4: Reviewer #1

No response.

Q4: Reviewer #2

There are several facets to this problem that should be considered. First, the examples listed in the question demonstrate that this parameter is not appropriate for all the target shorebird species because of differences among social systems. For those with polygamous or lekking social systems, determining the number of territorial males on a plot does not make sense. Secondly, the other examples listed in the question demonstrate that there are many other factors that make it difficult to determine the number of territorial males whose first nests are on a plot. The effects of these factors on the estimate of population size have to be considered in relation to (a) the estimate of number of birds on the intensive plots and (b) the estimate of number of birds on the rapid survey plots.

Because of the repeated visits and nest searching on the intensive plots, the estimate of territorial males should be close to the true number. However, without marked birds and intense observations throughout the egg-laying period, the estimate could be biased. The size of the bias is likely to be most influenced by predation rates. Because in arctic Alaska predation rates can have temporal trends over large geographic areas, this could induce a cyclical (and spatial) trend in bias. I don't know how significant this bias might be or what the implications would be for estimating changes in the size of populations.

For rapid survey plots, it is much more difficult to estimate the number of territorial males within a plot. With a single visit, the estimates will likely be highly inaccurate (resulting in high variance) but not necessarily biased in any particular direction. Again, bias will most likely be related to predation rates, which can be temporally or spatially biased. Reducing the variance among estimates of territorial males on rapid survey plots would likely have the greatest effect on increasing power to detect changes in population size. In this respect, there are several possible ways to reduce this component of variance.

First, the size of plot should be considered. With a single, rapid visit, one of the most difficult decisions to make is whether a bird observed on the plot actually "belongs" on that plot—i.e., is its nest or territory centroid actually on that plot? If this is the criterion to be judged, then fewer mistakes would be made the larger the plot is. Most of the errors have to deal with edges of the plots; thus, minimizing the edge-to-area ratio would reduce the likelihood of such errors. Increasing the size of a plot would also reduce the errors induced by birds coming in from adjacent areas to mob the observers. Judging whether birds belong on or off a plot can be particularly problematic in tundra areas for birds with large territories and low densities. Increasing plot size in these conditions would be helpful. This approach might also be helpful in situations where birds nest and feed in different habitats. If plots were configured to include both nesting and feeding territories, variance would also be reduced, not only because of fewer judgment errors on the part of the observers but also because the birds would not be missed while "away" on feeding territories.

For birds nesting in high densities, counting is more problematic. It is much easier to underestimate them than overestimate (which would likely happen more often with birds with large territories). Estimating numbers of territorial males is even more difficult with sexually monomorphic species nesting in high densities. Two- and three-bird chases, for example, can easily involve pairs and adjacent males or all males. Nests can be placed quite close together, so pairs can easily be missed.

One alternative to consider (and perhaps it has already been considered but not presented in these documents) is to use different criteria for counting birds during the rapid survey plots. It is not necessary, for the double sampling scheme, to have both surveys use the same methods. What is important is to have high correlation between the two. High variance in estimating birds using the rapid survey method will lead to poor correlation between estimates from intensive and rapid survey methods. Because the decision-making process of determining whether a bird is actually a territorial male on a plot is so difficult with a single, rapid survey, perhaps that should be simplified. For sexually monomorphic species, perhaps just a total number of adults observed on a plot would yield a more consistent index of breeding pairs. A "snapshot" approach might minimize difficulties that arise from movements of birds. This, coupled with increased plot size for areas where territories are larger, might reduce variability in estimates. Corrections would also need to be made for seasonal detectability of birds. Certainly as incubation progresses half of the population seems to disappear into the tundra.

Species with polygamous or lekking social systems are problematic with any of these approaches. Given that these species are also highly variable in their spatial distribution from year to year, it may be that no suitable means exists to monitor their populations adequately on the breeding grounds. Demographic monitoring for polygamous or lekking species in the arctic would be hopeless as well. I am curious as to how phalaropes were treated, since the methods as stated would not apply to their polyandrous social system.

Q4: Reviewer #3

The total number of birds on a plot is not being monitored!!! Rather the number of first nest territories is defined as the object of monitoring. I agree that a direct count of total numbers of birds would be difficult to interpret for some species, especially the ones that are characterized by (1)-(7) above. These are exactly the reasons why territorial males were chosen for monitoring. Even if not numerically the largest or the most obvious measure, nest territories are probably less variable in comparison with other potential population measures. Taking the number of nest territories and multiplying by 2 approximates a count of breeding adults, if sex ratio is equal. Or multiplying by 2.4 might include non-breeders if, for instance, another 20% more of each sex were known to be non-breeders on the breeding grounds. Or multiplying by 2.8 might include immature birds if another 20% were known to remain on wintering or staging areas. If derived from nest territories, an unbiased estimate of total number of birds in the population will depend on other data and other assumptions such as stable age structure. PRISM sampling intends to monitor the most important and probably the least variable component, the number of nesting territories. The true number of birds, or how the number male territories can best be re-scaled to approximate a total population size, is a different issue.

Q4:Reviewer #4

No response.

Q4:Reviewer #5

Again, answering these questions is extremely difficult without more information. For instance, it would be useful to know how many of the 22 species do not fit the assumed characteristics (i.e., territorial, monogamous, single-brooded, etc.), and how much they diverge.

The fundamental question, however, is whether the population size estimates based on making all these assumptions produce a number that is closely <u>correlated</u> with the actual number of birds that use the plot. For example, our research group recently looked into a similar question. We are studying a non-territorial, promiscuous species and wanted to know whether point counts were useful for monitoring. *A priori*, we guessed that point counts might not be especially good, since they largely detect singing males. Because of the breeding system in our species, this would not necessarily be a good indicator of the number of nesting females since females are not linked to a male or a territory. By conducting intensive banding on our study plots, however, we found that there was a relatively good correlation between the point count numbers and the number of birds banded (both males and females). Thus, we concluded that, for all their apparent imperfections, point counts can provide a useful index of population size (and thus trend). [As an aside, though we also found that point counts are lousy at predicting where most of the nesting occurs – so it would be a mistake to assess high quality habitat on the basis of point count surveys in this species.]

In order to address the questions raised, I believe that one would have to conduct similar studies with these shorebirds – i.e., test the hypothesis that the PRISM plot counts actually reflect the number of birds on a plot, as determined by some more intensive method (e.g., banding birds or nest searching). The work presented on detection ratios does this to some extent. For a monitoring program (where trend is really what one wants), however, I would be less concerned about determining the % detected, but rather would focus on the degree of correlation between the rapid and intensive population estimates – as long as these are well correlated the bias in the trend should be minimal. I have not been able to read all the supplemental documentation, so maybe these results have been presented elsewhere??

Q4:Reviewer #6

Population change will be confounded with detection uncertainties. I suppose that one could take a long view in that all things being equal, over-counts will balance under-counts. These sources of uncertainty will need further field studies to determine whether their effects are real.

Q4: SUMMARY OF REVIEWERS' COMMENTS

Two reviewers had no comments on this question. One reviewer agreed with the use of nesting territories as a parameter and felt that the method would deliver an unbiased estimate of the number of birds in a plot. The other reviewers judged that there could be bias resulting from various factors (predation, timing of breeding, and seasonal detectability were mentioned). These reviewers felt that further studies should be undertaken to determine if a) the PRISM counts truly reflect the number of birds on a plot; and b) if there is good correlation between what we is counted in intensive plots and what is counted in rapid plots.

Q4: RESPONSE TO REVIEWERS

We agree that variation in predation rates, timing of breeding, and seasonal detectability can all cause errors in the counts. It is important to recognize, however, that this only causes bias in the trend estimate if there is bias in the intensive surveys and the magnitude of the bias changes between survey periods. To avoid this bias, we strive for a complete count of nesting birds on the intensive plots. In 2004 we carried out a targeted research project to determine whether complete counts can be achieved.

Smith, P.A., Bart, J., Lanctot, R.B., McCaffery, B.J., and Brown, S. (2009) Detection probability of nests and implications for survey design. Condor, 111: 414-423.

*Abstract. Surveys based on double sampling include a correction for the probability of detection by assuming complete enumeration of birds in an intensively surveyed subsample of plots. To evaluate this assumption, we calculated the probability of detecting active shorebird nests by using information from observers who searched the same plots independently. Our results demonstrate that this probability varies substantially by species and stage of the nesting cycle but less by site or density of nests. Among the species we studied, the estimated single-visit probability of nest detection during the incubation period varied from 0.21 for the White-rumped Sandpiper (*Calidris fuscicollis*), the most difficult species to detect, to 0.64 for the Western Sandpiper (*Calidris mauri*), the most easily detected species, with a mean across species of 0.46. We used these detection probabilities to predict the fraction of persistent nests found over repeated nest searches. For a species with the mean value for detectability, the detection rate exceeded 0.85 after four visits. This level of nest detection was exceeded in only three visits for the Western Sandpiper, but six to nine visits were required for the White-rumped Sandpiper, depending on the type of survey employed. Our results suggest that the double-sampling method's requirement of nearly complete counts of birds in the intensively surveyed plots is likely to be met for birds with nests that survive over several visits of nest searching. Individuals with nests that fail quickly or individuals that do not breed can be detected with high probability only if territorial behavior is used to identify likely nesting pairs.*

We have tried to correct for timing/detectability issues by restricting rapid surveys to the period of time when the birds are establishing territories and laying eggs. At this stage they are 'tied' to their nesting area but flush more readily than later in incubation. We undertook studies on intensive plots in 2005 and 2006 to measure alternative prey (lemming and vole) populations in study areas and to understand the interaction between human activity and predation rates in plots. We now think that it is very difficult to measure lemming populations indirectly; a trapping survey needs to be instituted. We did not find any stable relationship between human activity and predation rate. Thus, the best way to minimize bias caused by predation will be to conduct frequent nest searches in intensive plots to ensure that nests are located before they are predated.

We have minimized one identified source of bias by changing the way that we count birds in a plot. We noticed that there was considerable variation in the way that surveyors estimated the number of territory centroids in a plot. Now, surveyors simply record the numbers of nests, probable nests, pairs and individual birds in a plot (excluding birds that are obviously not breeding; i.e., flocks). Surveyors are no longer expected to estimate number of territory centroids in a plot. The detection rates from the intensive plots are simply used as the correction. We judge that this new way of recording data will decrease our measurement error with respect territorial versus non-territorial birds.

The use of rope drags twice or more within the intensive surveys should minimize our risk of miscounting most territorial polygamous and polyandrous species. The only polygamous individuals that are potentially missed are those individuals who are holding territories but do not have nests. This scenario is rare to nonexistent in polyandrous species. Secretive species are also likely to be detected during the rope drags. It is also possible that some close-nesting individuals with ill-defined or no breeding territories (e.g. Red-necked phalaropes) could be confused and counted only once.

Re-nesting likely occurs to varying degrees, depending on predation pressure and weather in a given year. It is possible to quantify amount of re-nesting in a given area with intensive banding studies. However, the application of the results is limited temporally and spatially because of the great variation in nest success among locations and years. Again, frequent surveying in the intensive plots will detect likely re-nesting attempts, but will never provide a definite re-nesting figure without intensive banding.

The ‘problem’ of species with segregated feeding and nesting territories is no longer a problem if we are not assigning territory centroids for observations. The bird will be counted during rapid surveys, and the count is calibrated by results from the intensive plots (where the presence of feeding and breeding territories is incorporated into the detection rate). Similarly, species with large territories might be overestimated on rapid surveys if the survey plots occur on the edges of territories. Again, because we do not attempt to determine where the territory centroid lies, correcting rapid results with the detection rate should still yield an unbiased estimate.

Rapid surveys of intensive plots are repeated twice during the survey period, to account for variability in bird numbers in the plot at a given stage of the breeding season. As suggested by reviewer #5, we investigated the correlation between rapid count estimates of numbers in the intensive plots with the numbers that were actually breeding in the plots; the correlation is strong (R^2 = 0.62).

5. DERIVATION OF V($\hat{Y}$) AND $CV(\hat{Y})$

Is the derivation of V($\hat{Y}$) and $CV(\hat{Y})$ accurate? Can you suggest a better approach?

Q5: Reviewer #1

The authors appear to have gone out of their way to present their development in a confusing a manner as possible. First they write the detection ratio as $R = x/y$ rather than the standard development with $R = y/x$. Second they invert the use of the prime notation. (Standard development uses n for the intensive sample size and n' for the extensive sample) Third they fail to use the prime in a consistent manner the mean count of the rapid counts on the intensive plots is $\overline{x}'_h$ but the mean count for the exact counts is $\overline{y}_h$ on the intensive plots. The development could be clarified if they i) replace R with say P to denote detection probability, ii) point out clearly that they are have reversed the use of n and n' and iii) use $\overline{y}'_h$ to denote the exact counts on the intensive plots.

The presentation is extremely confusing. It appears that the terms camp, crew-year and primary sampling unit are all used interchangeably, showing a very cavalier attitude toward the mental health of the reader.

One arrives at equation 11 without any notion that the rapid counts will be replicated on the intensive plots but then in Table 3 the replicates are introduced into the variance equation. This leads to the question of whether it is appropriate to use an adjustment based on the average of two ($o' = 2$) rapid counts to adjust individual rapid counts

First sentence after equation (8) indicates that the rapid counts on the intensive plots will not be used in the average $\hat{\hat{X}}_h$. This is a reasonable approach if the intensive plots are not selected randomly. However, if intensive plots were selected randomly then this would be inefficient.

The expression for $V(\hat{\hat{X}}_h)$ (unnumbered equation between (9) and (10)) is based on assuming that the regression slopes β_{ha} are predetermined i.e. that the true slope of the regressions are known. This is a major assumption but we are given no hint as how the authors know what these slopes are. If the slopes are to be estimated from the data then the correct variance expression is much more complicated. Secondly the expression for $V(\hat{\hat{X}}_h)$ is missing terms for the correlations between $\overline{x}_h$ and the $\overline{u}_{ha}$. The variance expression should be

$$V(\hat{\bar{X}}_h) = V(\bar{x}_h) + \sum_a^E \beta_{ha}^2 V(\bar{u}_{ha}) - 2\sum_a^E \beta_{ha} Cov(\bar{x}_h, \bar{u}_{ha}) + \sum_a^E \sum_d^E \beta_{ha}\beta_{hd} Cov(\bar{u}_{ha}, \bar{u}_{hd}) \qquad \text{(BC 1)}$$

The error is apparent in the unnumbered expression by considering the case where E=1. In this situation the equation presented would have no covariance term.

Table 3 Section 3A is correct. Section 3B is more enigmatic, however, I think that estimator for $\hat{S}^2(\bar{x}'_{hkij})$ should be

$$\hat{S}^2(\bar{x}'_{hkij}) = s^2(\bar{x}'_{hkij}) - s^2(x'_{hijkl}) / o' \qquad \text{(BC 2)}$$

Table 4 The explanation in footnote 1 doesn't reflect what has been done to create the formulae here. First it is impossible to substitute Table 3C into Table 3B and then into equation (11). I believe what the authors want to say is substitute Table 3B into Table 3A and then into equation (11). However, this isn't what the authors have done. They appear to have managed to completely mess up the subscripts and confused the terms $\hat{S}^2(\bar{x}_{hk})$ with $s^2(\bar{x}_{hk})$ producing very confusing equations

For example the first entry in table 4 should be

$$g_{h1} = \frac{1}{\bar{x}_h^2}\left[s^2(\bar{x}_{hk}) + \sum_{a=1}^E b_{ha}^2 s(\bar{u}_{hak}) - 2\sum_{a=1}^E \sum_{d=1}^E b_{ha} b_{hd} \operatorname{cov}(\bar{u}_{hak}, \bar{u}_{hdk}) \right] \qquad \text{(BC 3)}$$

NOTE: THIS DOESN'T INCORPORATE THE FUNDAMENTAL ERROR DESCRIBED IN EQUATION (BC 1) ABOVE.

Perhaps it would be useful to introduce a Table 3D which gives combines 3A and 3B as

$$\begin{aligned} v(\bar{x}_h) &= \frac{(1-f_h)\hat{S}^2(\bar{x}_{hk})}{c_h} + \frac{\hat{S}^2(\bar{x}_{hki})}{c_h n_h} + \frac{\hat{S}(\bar{x}_{hkij})}{c_h n_h m_h} \\ &= \frac{(1-f_h)}{c_h}\left[s^2(\bar{x}_{hk}) - s^2(\bar{x}_{hki}) / n_h \right] + \frac{1}{c_h n_h}\left[s^2(\bar{x}_{hki}) - s^2(\bar{x}_{hkij}) / m_h \right] + \frac{s^2(\bar{x}_{hkij})}{c_h n_h m_h} \qquad \text{(BC 4)} \\ &= \frac{(1-f_h)}{c_h} s^2(\bar{x}_{hk}) + \frac{f_h}{c_h n_h} s^2(\bar{x}_{hki}) \end{aligned}$$

Note that the term in $s^2(\overline{x}_{hkij})$ drops out of the equation. This is because the among-plot variability is perfectly incorporated into the among cluster variability. (Because in this particular case the plot and cluster sampling fractions are negligible) This implies that the terms g_{h3} can be discarded.

Finally it must be pointed out that the theoretical development is based on assuming a large sample. For a ratio estimator the variance equation (10) is only appropriate when the number of intensive counts is a least 30 and the CV for the numerator and denominator are <0.1. Since the estimation is done separately within each stratum, these conditions must be applied per stratum. The total number of intensive counts is $c'n'm' = 45*2*4 = 360$ which must be divided among 51 strata. This indicates that there will be on average 7 intensive counts per stratum. This is far too small to justify the large sample theory. Further these average 7 counts won't even be a random sample for which the criteria of 30 was developed but are a multistage sample using plot in clusters and clusters in camps which further reduce the effective degrees of freedom.

The other criteria that the numerator and denominator have a CV<0.1 considered but given that the overall CV is only intended to be 0.31 it isn't likely that this standard is met.

Q5: Reviewer #2

I leave these formulas to the statisticians to verify

Q5: Reviewer #3

Although I readily admit my limited experience in statistical theory, the derivations appear appropriate to me.

Q5: Reviewer #4

I am not sure the regression on habitat variables will reduce the variance, because many more parameters are needed to estimate the bird habitat relationships. I suggest that the proposed method be compared with straight double sampling (see for example S. K. Thompson. 2002. Sampling. Wiley. Pages 158-159) for existing datasets.

Q5: Reviewer #5

I've left this question to the statisticians.

Q5: Reviewer #6

No response.

Q5: SUMMARY OF REVIEWERS' COMMENTS

Detailed comments were provided by only one of the reviewers, and he questioned several of the details (see responses below). Another reviewer suggested that we not use the habitat covariates.

Q5: RESPONSE TO REVIEWERS

We appreciate the detailed comments by reviewer 1 some of which we agree with and some we do not, as described below. The problem we encountered in the notation was that in the usual formulation for a ratio estimator, $r = y/x$, x is the true value. However, with other estimators, the parameter of interest is nearly always y (i.e., see any chapter in Cochran other than chapter 12 on double sampling). Thus, we had to choose between the usual formulation for ratios and defining the parameter of interest (population size) as x. Since population size was defined first, we thought it best to keep y=population size and invert the notation of the ratio. In an earlier draft we explained this, but decided that relatively few reviewers would be familiar with standard notation for survey sampling and that it might just confuse them. In the revision, however, we will re-insert the explanation for the notation. The reviewer says we used the prime notation inconsistently, implying an error in notation, but this is not so. We used $\overline{x}_h$ and $\overline{y}_h$ for all the rapid and intensive counts respectively and $\overline{x}_h^{'}$ as the subset of $\overline{x}_h$ counts that occurred on the intensive plots. We thought this would be clearer than defining $\overline{x}_h^{'}$ as the reviewer suggests. Given his suggestion, however, we will follow the notation he recommends in the revision. He suggests we replace R with P "to denote detection probability" but R is not a probability (i.e., it can exceed 1.0) so we would prefer to retain our notation.

The reviewer states that camp, crew-year, and primary sampling unit seem to be used interchangeably. Actually, however, the three terms mean different things and all three are needed. Camp refers to a location for an intensive crew (i.e., they do not move around). Crew-year refers to one rapid crew working for one year. Primary sampling unit refers to a crew-year when three-stage sampling is used but to a cluster of plots when two-stage sampling is used. The reviewer's comment, however, made it clear that we must explain these distinctions better. We have endeavored to clarify these three terms in the revised draft.

Re. eqn. 11, we made it clearer in the revised draft that multiple rapid surveys are usually made on the intensive plots (see Bart and Smith, submitted). The reviewer seems to question using the "average of the two rapid counts" but this is the plan we are investigating. We do not understand why this causes concern.

Bart, J., and P.A. Smith. (Submitted) Design of future surveys. Chapter 13 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

We agree with the reviewer's fourth paragraph about the intensive plots (and he does not recommend any changes).

The fifth paragraph raises what is apparently a major issue to the reviewer (he refers to it as a "fundamental error" later on). It involves replacing the estimated regression coefficients with the actual ones and thereby simplifying the variance calculation. The reviewer objects to this, but it is the procedure recommended by Cochran (1977, p. 194), who is certainly considered an authority on survey sampling, and we cited the page on which Cochran makes this recommendation and explains the rationale for it. Thus, we do not agree with the reviewer's criticism, nor do we think we needed to do anything more than refer the reader to Cochran for an explanation of the method. The reviewer also asserts that we have missed a covariance term, but here again we simply followed Cochran (1977, p. 195) who explains that this covariance term may be ignored because it is small compared to the other terms. We did not point this out but will do so in the revision.

The reviewer then criticizes our derivation of the g-values. His expression (BC 2) is correct for an estimator but we were not presenting an estimator, we were presenting a formulation for the true variance so we can investigate allocation of effort. There is a fundamental difference between these two tasks, especially in multiple-stage sampling in which the formulas for the variance and the estimated variance are different. We believe that the reviewer did not follow our rationale. His comment shows that we need to be clearer in describing our approach but it does not identify any error in the expressions we used.

The problems above account for the next paragraph and the reviewer's expression (BC 3). This material was derived by J. Bart who has reviewed the notation, subscripts, and expressions and believes they are correct. In deriving them, he wrote out extremely detailed explanations (available upon request) and, while we considered sending them to all the reviewers we doubted (incorrectly as it turns out) that anyone would want to wade though them. In retrospect, we wish we had provided this extensive material as it would probably have helped the reviewer understand our rationale.

The difference between what the reviewer seems to be doing, and what we did, is evident, in his comment that g_{h3} can be discarded because it is "perfectly incorporated into the among cluster variability". This is true *in estimation*. But to investigate allocation of effort we had to obtain estimates of each variance component. The term, g_{h3}, refers to variation in rapid survey results among plots within a cluster.

Ignoring it would be equivalent to claiming that precision is unaffected by how many plots one does within a cluster. But obviously precision will be higher if more plots are surveyed per cluster. Thus, it cannot be ignored (i.e., "discarded").

The reviewer's last comment on this question pertains to use of large sample theory for the variance estimation. He points out – correctly – that we implied estimation of separate ratios in each stratum, and notes that large sample theory does not then apply. This is correct. We did not mean to imply that separate detection ratios would be estimated for each stratum and, as the reviewer points out, this would not even be possible. Detection ratios are combined across strata, with among-strata variance in detection rates contributing to the variance of the overall rate. We have corrected the estimation equation accordingly and appreciate the reviewer pointing out this error.

The manuscript has been re-written to incorporate all of the changes identified in these notes, but also to make an additional change in the estimation equations (to acknowledge lack of independence between strata). The new estimation approach has been provided to this reviewer who has agreed to provide comments on it. We agree with reviewer #4 that the habitat models appear to add little and may even be reducing precision (by adding parameters). In the revision, we have omitted this step. These changes necessitated a reanalysis of the data, which has been completed and incorporated into the revised manuscript. However, the results of our power analysis did not change substantially as a result of this reanalysis, and the conclusions which were drawn from the previous analysis remained largely unchanged (Bart et al. *submitted*, Bart and Smith *submitted*).

Bart, J., V. Johnston, P.A. Smith, A. Manning, J. Rausch, and S. Brown. (Submitted) Methods used in the Arctic PRISM surveys. Chapter 2 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

Bart, J., and P.A. Smith. (Submitted) Design of future surveys. Chapter 13 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

6. STRATIFICATION

Does the proposed stratification system appropriately define region-habitat strata that are relevant for the distribution of breeding shorebirds in the Arctic and adequately characterize species densities across the landscape? If not, can you suggest a better stratification system that should be followed?

Q6: Reviewer #1

No response.

Q6: Reviewer #2

I found this section extremely confusing and some of the assumptions questionable. Most of these stemmed from the perceived need (with which I disagree) to estimate total population size. The assumption that I find most tenuous is that the ratio of densities in wetlands and other habitats holds throughout a species' range.

Q6: Reviewer #3

A definite strength of the proposed PRISM sampling method is that a complete and consistent geographic framework is used for the total of all 51 "geographic region-CAVM habitat" stratum areas to be sampled. For straight-forward comparison of total numbers over time, the sampling framework defined by these 51 "geographic region-habitats" must be held constant. The proposed number, size, and boundary delineations should be carefully re-examined for all sorts of practical considerations by field biologists experienced in arctic field camp logistics, GIS practitioners, and data analysts. Changes should be made before they become fixed by a large amount of expensive sampling data. Although a region-habitat stratum can probably be subdivided, provided that adequate sample locations exist in each subdivision, they cannot easily be combined without potential bias due to different historic sampling intensity.

I have to admit some confusion as to the derivation of the 51 habitat-regional strata because I am not familiar with the accuracy or the scale of the CAVM boundaries, other than what is seen on the map. It seems to me that the sampling becomes locked to these areas, which is OK, especially because it is likely that no better data exists at his time. Nevertheless, I would question whether some of the vegetation types, often those with a lower density of shorebirds (shrub, sparse herbaceous, etc) should be kept in separate sampling strata. Some combination has already been done but perhaps more combinations would save some amount of sampling effort and still provide useful estimates. Perhaps the more detailed vegetation types should be used in the lower-level stratification or the regression models for each species. But the terminology has confused me, and I'm not very clear on how the "upper" and "lower" habitat information fits together.

Also there was a mention in the text (p10) that indicates the size of these areas will in fact be estimated because large lakes, % water, non-vegetated barrens, and spring flood areas are to be subtracted from the sampling framework. This seems a little tenuous. It might be preferable to use total geographic area as a known constant for each stratum, and then define some map GIS coverage, satellite remote sensing data, or visual reconnaissance procedure to estimate the amount of non-nesting habitat and best delineate its exact boundary. In my experience, because of map location error, classification error, and habitat change over time, the exact definition of non-nesting-habitat may not be a trivial problem. If it can be 100% defined, this non-habitat will not need to be sampled (or sampled at a very low rate) because the bird density is 0. Alternatively definition of non-nesting habitat (e.g. water, barrens, and snow fields) could be part of the regression model.

Combining the last 2 comments, perhaps each geographic region should be stratified into high, low, or none (crude estimate of shorebird density) based on some combination of pilot studies, reconnaissance, and imagery. These boundaries once established would be held constant. Hopefully less than 51 strata could be established.

The accuracy of the fine scale habitat classification within each stratum, or how the plots are classified based on characteristics (elevation, hydrology, satellite TM, or whatever combination of data), will not cause bias in the estimated mean stratum density. However the habitat modeling and establishing regression model relationships are critical for increasing the precision of the estimate within each region. As I understand it, the fine-scale habitat classification can change with time as better data become available, and the habitat classification may differ among regions. The random selection of plot-clusters must not be conditioned on these fine-scale habitat or regression variables. Although I am not entirely sure, it seems reasonable that the sampling intensity (probability of selection of secondary units, the rapid plots) could be adjusted and allocated provided that the same regression model data were available throughout all the sampled plot-cluster area if not the entire stratum area.

Q6: Reviewer #4

Fifty one is a large number of strata. One loses one degree of freedom (1 site) for each stratum, so we should consider whether or not that degree of stratification will reduce the variance enough to compensate for the loss of degrees of freedom. I suggest that a much smaller number of strata would be appropriate. If stratification is being used to equally distribute the sample, consider using systematic sampling.

Q6: Reviewer #5

I’ve left this question to the Arctic biologists.

Q6: Reviewer #6

From my experience, your strata do partition the arctic breeding areas adequately. Of course, within these there are abrupt local changes that do affect shorebird numbers and composition: e.g., in Region 12 (from Fig 1) the Anderson River Delta is especially rich while the areas extending east and west are not.

Q6: SUMMARY OF REVIEWERS' COMMENTS

Two reviewers did not comment on this question. One felt that the proposed stratification is appropriate for arctic shorebird surveys. Two suggested that the number of strata should be reduced to make sampling effort more efficient. One doubted that the assumption that the ratio of bird densities in wetlands and other habitats would be constant throughout a species' range is correct. Two found the explanation of habitat stratification confusing.

Q6: RESPONSE TO REVIEWERS

If the habitat strata capture variance in the density of shorebirds, the precision of the estimates is improved. However, as the number of strata increases, the degrees of freedom for the test is reduced for a given sample size of plots or clusters. There is therefore an optimal level of stratification. Although we cannot make the claim that we have achieved this optimum, results to date do suggest that our stratification captures variation in shorebird densities. More importantly, our power analyses suggest that we will meet the accuracy target for most species with reasonable levels of effort.

Further, we now recognize that some flexibility in rules for stratification is important. For example, an oil and gas pipeline has been proposed in the Mackenzie Delta, one of the most important breeding and migration sites for birds in Arctic Canada. This proposal has led to detailed surveys of this area during the past few years, especially at the proposed gas production facilities and on the pipeline corridor. During PRISM surveys, the Mackenzie region was sub-divided into more sub-regions than would normally have been the case for an area this size, and the pipeline corridor and proposed facilities were delineated as one small, non-contiguous sub-region. This approach permitted biologists to conduct a higher number of surveys in the areas proposed for development than in other parts of the region, yet the results could both be integrated with, and compared to, results from surrounding sub-regions.

For additional details, see:

Bart, J., V. Johnston, J. Rausch, P.A. Smith, and B.J. McCaffery. (Submitted) Priorities for future PRISM surveys. Chapter 15 in J. Bart and V. Johnston (eds.). Shorebirds in the

North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

As per Reviewer #1's suggestion, the authors undertook a final close examination of the region-habitat strata boundaries, to ensure that they are as accurate as possible. We made minor adjustments to the region boundaries. As each region is surveyed (reconnaissance and first round of PRISM surveys), we expect that further adjustments will be made.

One reviewer questioned the constancy of wetland distribution. We agree that it is variable, but this only affects precision (it causes no bias) in the population estimate. The reviewer was concerned in particular that the ratio of birds in wetlands versus uplands might differ, and that this would complicate our power analysis. We carried out a revised power analysis with substantially more data; the wetlands:uplands ratios were no longer used.

The section of the manuscript describing our habitat stratification method has been rewritten as follows:

Bart, J., V. Johnston, P.A. Smith, A. Manning, J. Rausch, and S. Brown. (Submitted) Methods used in the Arctic PRISM surveys. Chapter 2 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

> *The study area is the Arctic portion of North America, as delineated on the Circumpolar Arctic Vegetation Map (CAVM; CAVM Team 2003), with modifications to exclude some mountainous areas (Fig. 1). The study area was partitioned into 19 regions (Fig. 1). Prior to conducting surveys in each region, the region was further divided into sub-regions on the basis of how much area could be covered by field crews, and anticipated density of shorebirds. Large areas not suitable for nesting shorebirds, such as oceans, lakes, and barren areas, were excluded.*
>
> *Each sub-region was partitioned into plots, most of which covered 0.12 - 0.16 km*^{2}*. Plots were substantially larger in some early years of the study. We defined "wetland", "moist", and "upland" habitats in each sub-region and calculated the proportion of each plot covered by each habitat type. Plots with no habitat suitable for nesting birds were deleted. Plots with only small amounts of wetland, moist, or upland habitat that were primarily covered by water caused various logistic and statistical problems and were therefore combined with surrounding plots…*
>
> *Plots were assigned to wetland, moist, and upland habitat types. The rules used to make these assignments varied across the Arctic because the extent of different habitats varied substantially. In many cases, plots were assigned to the type corresponding to the habitat that covered the largest fraction of the plot*

(e.g. if wetland habitat covered >50% of the plot, the plot was assigned to the wetland plot type). If wetland habitat was rare within a sub-region, a different rule was used, e.g. "if wetlands habitat covers >20% of the plot, then assign the plot to the wetland plot type; otherwise, assign the plot to the type corresponding to the habitat that covers the largest fraction of the plot". The rules used in each region are described in more detail in the regional reports (Chapters 3-8).

The sampling plan for selecting plots to survey involved stratification using sub-region and plot type (wetland, moist, upland), followed by selection of clusters of plots and then selection of plots. We selected plots to survey in groups to reduce distances between plots being surveyed at the same time. These groups usually included plots in different habitats, and thus in different strata. Selection of plots was thus not independent in different strata (i.e., plots in different strata were close together much more often than if we had used independent selection). We referred to the groups of plots as "zones" to distinguish them from clusters which, by definition, are plots in the same strata. Most zones covered 4 - 36 km^2 and comprised 25 to a few hundred plots. We acknowledged the lack of independence, in selecting plots in different strata, by modifying the standard formulas for cluster sampling (see below). Some reviewers have had difficulty grasping why we had to define zones but had we ignored the lack of independence, caused by selecting plots in different strata within zones, our variance estimates would have had substantial negative bias. Zones to be surveyed were selected systematically to ensure even coverage across the sub-region. Simple random sampling was used to select plots within clusters.

In the early years of the study, we attempted to carry out the steps above by hand. With large sub-regions, this was not possible and we were forced to use short-cut methods which inevitably caused us problems later in the analysis. We therefore prepared a series of ArcGIS tools, collectively referred to as the Arctic PRISM ArcGIS extension (Table 1), to automate delineation of plots and assignment of plots to clusters, zones, and strata. This tool was essential for partitioning large regions into plots. It is available free from the senior author and may be useful to others who need to define a rigorous sampling frame for large, heterogeneous areas.

7. COVERAGE OF RARE SPECIES

Will the low level of sampling in some very large strata likely miss rare or locally distributed species? If yes, can the proposed plan be modified to correct for this problem? If the problem cannot be corrected, how will this fact affect the ability to detect population trends for these species?

Q7: Reviewer #1

If the population is rare or locally distributed then the variance estimates used in the design are underestimated. A well designed survey with random sampling is the best scientific approach available. However, being aware of the limitation of sampling schemes in which a vanishing small fraction of the surface area is actually visited is important. The scheme to collect anecdotal and happenchance information on species distributions which would help identify the existence of local concentrations would be useful for an improved design in the future.

Q7: Reviewer #2

Certain species, such as Black Turnstones, have spatial patterns of distribution that are poorly sampled by this protocol. Because their patterns don't shift much interannually, however, their population size could be monitored better by sampling replicated plots over time. An adequate number of plots would need to be monitored within their core breeding area to achieve the desired power. For other species, such as Buff-breasted Sandpipers, which occur in low numbers but undergo dramatic interannual shifts in distribution, such a protocol may never achieve the desired power to monitor population changes.

Q7: Reviewer #3

The chance of missing rare or local species is not any more a problem than the problem of oversampling rare or local species. In fact it is perhaps harder to guard against the oversampling problem because of the natural tendency to high-grade areas and pick samples where better habitat and more birds are thought to occur. A systematic sample, or randomization restricted to guarantee some minimum spacing between plot clusters, might be considered.

Q7: Reviewer #4

An important reason for stratification is to assure that there is an adequate sample size in less common habitats.

Q7: Reviewer #5

Yes – it is hard to see how it could not. Any rare or locally distributed species is going to be difficult to survey with any randomized scheme with limited sample size. Within the framework of PRISM my guess is that the only ways to solve this problem are (a) to increase the sample size a lot, or (b) to add in non-random samples that are explicitly intended to increase the probability of encountering rare or local species. The first solution, obviously, is hampered by logistical constraints, while the second creates all kinds of problems in the interpretation of the data. I would suggest that PRISM, or any PRISM-like program, is really not suitable for rare or local species, in just the same way that the BBS is no way to monitor California condors or Kirtland's warblers. Rare and local species can only really be monitored adequately through focused surveys directed at the individual species.

Q7: Reviewer #6

I suspect that this will be a problem. See below on permanent plots sampled annually.

Q7: SUMMARY OF REVIEWERS' COMMENTS

Reviewers agreed that achieving the power objective will be difficult for rare species. One reviewer stated that the variances are underestimated for rare species.

Q7: RESPONSE TO REVIEWERS

As noted under "survey interval" we agreed that some aspects of the power analysis required revision and we therefore conducted a new analysis. See response to question #3, "Survey Interval" for a summary of the results.

This analysis suggested that the accuracy target would be met for 24 or 25 of the 26 species. However, it was also evident that the species that have proven most difficult to survey in the Arctic are relatively rare or have a restricted distribution. The issue of which species are easily monitored by the arctic PRISM methodology, versus those species for which additional surveys might be beneficial, is explored in:

Skagen, S.K., P.A. Smith, B. Andres, G. Donaldson and S. Brown. (Submitted) Contribution of Arctic PRISM to Monitoring Western Hemispheric Shorebirds. Chapter 16 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

> *While rarity can make a species difficult to survey throughout the annual cycle, a restricted distribution may in some cases be an asset for targeted, single-species surveys. High fidelity of Pacific Golden-Plovers to their Pacific island wintering site suggests that information on changes in their population size could be obtained there (Johnson et al. 2006). Virtually all of the Hudsonian Godwits wintering along the Pacific Coast do so in the vicinity of Chiloé Island, Chile (Andres et al. 2009), and systematic ground counts could provide information on population size and trends. Recent analysis of Christmas Bird Count data (Butcher and Niven 2007) may prove useful for tracking changes in a select group of shorebirds that winter in North America, such as the Purple Sandpiper. A thorough review of alternative methods for the species not surveyed well by Arctic PRISM should be undertaken.*

We do not know why one reviewer said the variances are under-estimated for rare species and do not agree with this comment.

8. ESTIMATION OF G-VALUES

Do you agree with the approach of using models to estimate the g-values rather than using species-specific values? Are the assumptions implicit in these models reasonable? If not, can you suggest a better approach? (Note: the models were used for the power analysis but species-specific values may be used when more data have been collected.)

Q8: Reviewer #1

The use of models to estimate g-values is a practical approach to designing a survey with limited data. It would be useful to review the predicted g-values once sufficient data has been collected to derive reliable species specific estimates. Since the g-values are critical to assessment of the adequacy of the design it might be prudent to use overestimates to build in a safety margin.

Q8: Reviewer #2

I am skeptical about this approach, but it would have been easier to evaluate if the values in the plots had shown individual species. Perhaps I am wrong, but it seems that this approach is likely to underestimate the variance that would actually be recorded in a given set of surveys. It seems that a better approach would have been to do simulation modeling with actual values. It is true that variance is likely to change as population size changes and that this should be taken into account when conducting power analyses. In three of the graphs in Fig. 2 the equation did not match the labeling of the axes—I was not sure if these were on a linear or logarithmic scale.

Q8: Reviewer #3

No response.

Q8: Reviewer #4

See #5.

Q8: Reviewer #5

I'll leave this question to the statisticians too.

Q8: Reviewer #6

No response.

Q8: SUMMARY OF REVIEWERS' COMMENTS

Only two reviewers responded to this question, one seemed satisfied with it while the other expressed some skepticism suggesting that the variance might be underestimated (but not saying why).

Q8: RESPONSE TO REVIEWERS

With the much larger data set available for analysis now, we no longer used the modeling approach referred to in this question. In estimating density and population size we did not extrapolate outside the surveyed regions, so we always had empirical data. For the power analysis, where we had to have estimated values for all regions, we used neighboring regions as "surrogates" rather than a modeling approach. While this approach may still over- or underestimate densities in some unsampled regions, we feel that it is a much more defensible approach, and wholly adequate for an assessment of power.

9. EVALUATION OF SAMPLING PLANS

Is the approach used to evaluate candidate sampling plans appropriate? If not, can you suggest a better approach?

Q9: Reviewer #1

The approach sounds reasonable, however, allocating samples to habitats as fraction of the total is risky concept. The allocated sample sizes may not be integers and may result in unrealistic or impractical sample allocations. The results of the allocation should be written out explicitly by habitat to ensure they are practical.

Q9: Reviewer #2

This seems like a reasonable approach for determining the optimal sampling plan to reach the desired CV for the most species.

Q9: Reviewer #3

No response.

Q9: Reviewer #4

Simulation is certainly very appropriate with the complexity of this estimation. However, in #5 I suggest using a simple double sampling approach, which could be evaluated analytically.

Q9: Reviewer #5

It made sense to me. Though I'd like to see the economic trade-offs addressed explicitly (more on this later).

Q9: Reviewer #6

No response.

Q9: SUMMARY OF REVIEWERS' COMMENTS

The reviewers generally seemed satisfied with the approach although one commented that strict adherence to proportional allocation can lead to problems.

Q9: RESPONSE TO REVIEWERS

We agree with the problem about proportional allocation among habitats, and use the target ratios of plots in each habitat type as a guideline only (e.g. Smith et al. *submitted*). For the remainder of the optimal allocation exercise, we feel that this level of detail is adequate for assessment of feasibility at this stage. As surveys are carried out, and more data become available, it will be possible to revisit the issue and determine whether our predictions about optimal allocation were correct.

Smith, P.A., V. Johnston, and J. Rausch. (Submitted) Southampton and Coats Islands. Chapter 6 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

10. ESTIMATION OF POPULATION SIZE

The estimates of current population sizes, and the related assumptions concerning densities between regions and habitats within regions, play a very important role in the overall analyses. Is the use of this region-habitat stratification system to estimate population size (see ms. pages 10-11) appropriate? If not, can you suggest a better approach?

This section compares the population estimates provided by Morrison et al. (2000) with estimates derived from field data. Various factors were considered to develop the single population estimate for each species. Are all final population estimates reasonable for the power analysis? If not, indicate the inappropriate estimates and justify the use of another population estimate.

Q10: Reviewer #1

No response.

Q10: Reviewer #2

I think that this section is on very shaky ground. First of all, the maps (Wildspace 2002) used to determine the breeding ranges are grossly inaccurate for many of the species in Alaska. I am not surprised, therefore, at the incredible discrepancies between the PRISM estimates and those from Morrison et al. (2001), which were off by a factor of up to 800%! There should have been a careful effort to determine from the primary scientific literature the best information to use, not only for breeding ranges, but also for population estimates. Much more accurate range maps are available from the *Birds of North America* accounts. For Black Turnstones, the population estimate of 80,000 birds is only for those breeding on the central Yukon-Kuskokwim Delta, where surveys were conducted; the total population was estimated at 95,000 range-wide. This information can be found in the original published paper.

I think that the whole process of determining whether the PRISM estimates were reasonable or not was fairly arbitrary and not carefully evaluated. Again, I question whether this approach is really the best, even given the best data available on breeding distribution.

A couple of other notes: (a) headings for Tables 7 and 8 are confusing since they seem to be reporting the same things (after careful scrutiny I could figure out how they differed); (b) riparian is listed as a habitat type in Table 7 but not mentioned anywhere else in the methods or results. How was this really treated in the analysis?

Q10: Reviewer #3

No response.

Q10: Reviewer #4

No response.

Q10: Reviewer #5

My biggest concern in this section was that I felt it lacked detail (and justification) for the ways in which population sizes were adjusted to accord with Morrison's estimates. For instance, estimated densities and thus the overall population estimate for American golden-plover was adjusted downwards to bring it more in line with Morrison (though it is still 8 times greater), but it is not clear why/how the adjustment decisions were made. For other species there is even less detail: e.g., "we adjusted our estimates until they produced the Morrison et al. estimate" for Pacific golden-plover is often the extent of the explanation.

For most species, the PRISM surveys generated estimates that are far higher than have been made previously. This type of result is perhaps not unexpected given that extrapolations are being made over large areas, and that the GIS data involve grouping habitat into very broad types. I wonder if error in the GIS (i.e., over-predicting the area of certain habitats by not excluding unsuitable sub-habitats) might not be a major issue here. Given the large errors that could be generated by multiplying over large areas it seems as though some detailed ground-truthing of the habitat data should be a high priority. Another possible cause of the discrepancies could be over-counting of birds on the PRISM surveys – this could conceivably be tested by separate studies that use marked birds to estimate the extent to which miscounting occurs when using the PRISM protocol. Of course it could be that earlier estimates were just way off, but some of the differences are very large. It would be nice to see how the difference between the PRISM and Morrison estimates varies as a function of Morrison's accuracy estimates. One would hope that the discrepancies are positively (and systematically) correlated with Morrison's uncertainties.

I also would have liked to see confidence intervals placed around the population estimates – for all I know such an interval would be so wide as to encompass the Morrison estimates, in which case the downward adjustment might not be justified. Given the rigor of the design, this truly critical component of the approach – where population sizes are actually generated – seems rather *ad hoc*. Moreover, the authors' frequent decision to adjust their figures suggests a lack of confidence in the approach on their part, and is a sign that adjustments in the protocol might be warranted (e.g., to account for habitat ground-truthing). In general, I think it is important to generate better justified methods for adjusting population sizes – or just go with what PRISM produces (otherwise, what's the advantage of all the statistical rigor?).

Lastly, though, I should return to an earlier point: that for monitoring it might not matter whether these population estimates are off-base, as long as there is a well-supported, and consistent, relationship between the PRISM numbers and the true numbers. Thus, this section does not lead to me believe that PRISM is not useful for tracking trends. But I remain to be convinced that it can be used to estimate population sizes.

Q10: Reviewer #6

Yes it is. Stratification is necessary and your stratification is probably the best balance between practicality and scale (i.e., there are smaller scale entities but including them as strata would create too much complexity for the available effort).

Q10: SUMMARY OF REVIEWERS' COMMENTS

Comments were provided by only three reviewers, and only one addressed the issue of the use of region-habitat stratification for the initial estimates of population size. This reviewer felt that the method was appropriate. Two reviewers were concerned about the accuracy of the population estimates and the objectivity of the methods used to arrive at them. These reviewers suggested that more accurate range maps are available than those used by us, and that the large discrepancies between our estimates and previous ones may stem from these and other GIS inaccuracies.

Q10: RESPONSE TO REVIEWERS

This question is no longer relevant. In lieu of using range maps and published information to estimate population size for the purpose of power analysis, we used actual survey data. This approach was possible because of the large quantity of PRISM data obtained since the original power analysis was conducted. Numbers in surveyed regions were determined from observed densities, and densities from neighboring regions were used to represent unsurveyed regions. Data were available for a substantial fraction of the Arctic, so although our projected densities for unsampled regions may be slightly over- or underestimated, we feel that the errors would not be so large as to change the results of our power analysis significantly. Similarly, even substantial uncertainty about range is unlikely to have serious bearing on the conclusions of the power analysis.

11. ESTIMATION OF DETECTION RATES

A standard detection value is used for all species in these analyses. Does this concern you and, if so, can you suggest a better way to conduct the power analysis given the high SEs for species-specific detection rates.

Q11: Reviewer #1

The use of a single detection rate for all species is an unfortunate necessity given the lack of specific data. I would suspect that detection rates vary both by species and habitat. Would it be possible to model habitat by species detection rates in a manner similar to that used for g-values?

Q11: Reviewer #2

I think that it is absolutely incorrect to use a single detection value for all species in the analysis. The SEs of the detection ratios could probably be reduced for most species by implementing some of the changes suggested above in conducting rapid surveys, but this won't help for conducting power analyses now. An alternative would be to use a bootstrap approach, in which one value for detection ratio is randomly selected (with replacement) for each of the bootstrap samples from among all values gathered in the field. This should be done separately for each species. It is also quite probable that detection ratios depend upon density, so intensive sites should encompass that variability. I think that using a single detection ratio for all species, with its smaller SE, not only will provide inaccurate population estimates but also will underestimate variance of those estimates. This will result in erroneous conclusions about the power of this program to meet the monitoring goal.

Q11: Reviewer #3

This is not a problem for the power analysis.

Q11: Reviewer #4

I am not familiar with these species, but it would seem that detection rates should very among species. I am not surprised that precision was poor. As indicated above, I suggest that you look at double sampling without the habitat regression, reducing the number of parameters as a possible approach to improving the precision.

Q11: Reviewer #5

Since the decision to use a single detection value was based on the empirical observation that detection values for different species could not be distinguished, it seems to me to be clearly appropriate. Using different values would only be justified if there were good evidence that values really differed. That said, the lack of any difference doesn't mean that there is no difference, and it would be great if precision could be improved. Unless that can be done, however, I see no alternative to the approach taken.

Q11: Reviewer #6

You are attempting to survey a suite of species at 'one go' and for the sake of field logistics some compromises must be made. Furthermore, the high SE's will preclude any realized advantage derived from micro-designing this whole effort.

Q11: SUMMARY OF REVIEWERS' COMMENTS

One reviewer objected to the use of a single detection rate; the others either indicated that there was no other alternative (although they suggested evaluating this more closely) or supported the practice.

Q11: RESPONSE TO REVIEWERS

We now have considerably more data than when the peer review was started. In the monograph (Bart and Smith *submitted*), we generated species specific detection ratios for all species with adequate data. For most species, the 95% confidence interval included the combined rate. Whenever the SE of a species-specific rate was reasonably small (i.e., CV<0.3), we discussed what the effect of using the species-specific rates would be. A significant difference in detection rate was observed between Canada and Alaska (because of a difference in counting methods), so we used a region-specific combined rate.

Because the species-specific rates are reported, and the effect of using them is described explicitly, readers can evaluate the implications for themselves. Furthermore, once the monograph is published, all of the data and the programs to analyze it will be made available on a public web site so that interested users can analyze the data any way they want to.

We considered the issue of larger variance for species-specific rates of detection carefully, and our analyses in the monograph show that variances of species-specific rates are not uniformly larger, as one reviewer predicted (Bart and Smith *submitted*). For common species, the species-specific rate often had a lower SE than the combined rates. For rare species, however, the species-specific SE was usually larger, and sometimes much larger.

For additional details, see:

Bart, J., and P.A. Smith. (Submitted) Summary and conclusions. Chapter 14 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

12. VARIATION IN ESTIMATED DETECTION RATES

Initial efforts to generate species-specific detection rates have produced estimates with considerable variation across species due to differences among regions and individual surveyors. Does this concern you, and if so, do you think this fact undermines the use of a double sampling approach to monitor arctic-nesting shorebirds which depends on using a detection ratio as a correction factor?

Q12: Reviewer #1

In general I am unhappy with the general concept of inflating observed counts to population levels by simply dividing by the detection rate. Although there are several papers and texts which state that this is the "natural" approach to working with detection rates, there are several important ideas which have been glossed over. The main problem is that even if the estimate of the detection probability (p) is essentially unbiased, the reciprocal of the estimate (1/p) will not be unbiased and may in fact have a substantial bias. The conditions when (1/p) will be relatively unbiased are when p is large and it has a small variance.

The ratio estimator is based on the unrealistic model that the E(y|x)= bx i.e. that the relationship of the true count to the actual count is a regression line through origin which can be paraphrased as saying that when no birds have been recorded then there were none there to be recorded. A more effective model for the relationship is to use a regression estimator. For example assume the actual counts have an underlying negative binomial distribution.

$$f_y(y) = \frac{\Gamma(r+y)}{\Gamma(r)y!} p^r (1-p)^y \qquad \text{if } y>0$$

and that using the rapid count each bird is detected independently with probability p. Hence the rapid count given the true count has a binomial distribution

$$f_{x|y}(x \mid y) = \binom{y}{x} t^x (1-t)^{y-x}$$

Hence the joint distribution of the 2 counts is

$$f_{x,y}(x,y) = \frac{\Gamma(r+y)}{\Gamma(r)x!(y-x)!} p^r (1-p)^y t^x (1-t)^{y-x}$$

and the marginal distribution x becomes

$$f_x(x) = \frac{\Gamma(r+x)}{\Gamma(r)x!} \dot{p}^r (1-\dot{p})^x$$

where

$$\dot{p} = \frac{p}{p + t(1-p)}$$

which is also a negative binomial distribution.

Finally, the distribution of the true count given the rapid count is given by

$$f_{y|x}(y \mid x) = \frac{\Gamma(r+y)}{\Gamma(r+x)(y-x)!}\ddot{p}^{r+x}(1-\ddot{p})^{y-x}$$

where

$$\ddot{p} = p / \dot{p} = 1-(1-p)(1-t)$$

Thus for x>0, the distribution of y given x has a negative binomial distribution displaced by x giving

$$E(y \mid x) = r\frac{(1-\ddot{p})}{\ddot{p}} + \frac{1}{\ddot{p}}x$$
$$= \beta_0 + \beta_1 x$$

For case of shorebird nesting data the population density is extremely small and the difference between assuming a regression line through origin or not could have a substantial influence on the bias of the estimator.

Q12: Reviewer #2

I applaud the effort to use a double-sampling approach and encourage efforts necessary to generate species-specific detection ratios with small standard errors. I think there are ways to adjust the protocol to reduce the inconsistency among observers, as I've outlined above.

Q12: Reviewer #3

The variation among species, regions, and observers in detection rate is a problem that may limit the final precision of density estimates. I think the same problems would exist if other field methods were used to estimate density, although many methods would not detect that the variation existed. Masking or avoiding a straight-forward estimate of detection ratio will not solve the problem. With appropriate training, experience, and attention to possible covariates such a weather and timing, the variation can be minimized.

Q12: Reviewer #4

Double sampling is a sound approach and should not be abandoned although it could be simplified as suggested above.

Q12: Reviewer #5

There is clearly a concern here, as the results suggest additional sources of sampling error due to observers and regions. This error (I believe) should not create bias, however, only a lack of precision. So, the concern centers on the need for a larger sample size to detect a trend than would be the case if detection rates were less variable. The lower power that this result implies, further argues for the need to use a more liberal α-level (see question #1).

It would be good to know how the variation is apportioned between regions and individual surveyors. If most of it is due to the latter, then it might be something that PRISM is stuck with – some regions are inherently harder to detect birds in than others. If, however, much of the variation is attributable to observers, then it might be something that can be improved, e.g. by better training, ensuring continuity in field crews from year to year, etc. This might be worth looking into.

Q12: Reviewer #6

No response.

Q12: SUMMARY OF REVIEWERS' COMMENTS

One reviewer acknowledged that our approach is widely accepted but suggested ways to improve on it. The reviewers generally recognized that variation in detection rates decreases precision but does not cause bias.

Q12: RESPONSE TO REVIEWERS

We are intrigued by the reviewer's suggestions for a new approach to estimating detection rates but have not had the opportunity to explore it. Also, we acknowledge a mistake in the original manuscript where it was suggested that detection rates would be calculated across species for each stratum. As the reviewer points out, this is impractical, and it was never our intention. But the manuscript was confusing on this point and it has been re-written to make this clearer (Bart et al. *submitted*). By using a "ratio of the means" approach, rather than a "mean of the ratios" approach, the bias becomes negligible (e.g. Cochran 1977, Chapter 6).

Bart, J., V. Johnston, P.A. Smith, A. Manning, J. Rausch, and S. Brown. (Submitted) Methods used in the Arctic PRISM surveys. Chapter 2 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

13. ALLOCATION OF EFFORT

Do you feel that the section (3.4) on allocation of effort supports the conclusion at the end of the section?

Q13: Reviewer #1

In general allocation of effort between samples and sub-samples is fairly robust procedure with comparable results over a wide range of allocations. The results in section 3.4 seem reasonable.

The approach described in section 3.5 seems very inefficient approach to the problem but is plausible. However, the results given in Table 12C are very counterintuitive. I would anticipate that allocation to a stratum should increase as the number of crew-years increases. However there are several instances in which the crew years per region decreases as the total number of crew years increases.

The allocation is based on the 15 amalgamated regions(not the original 19 regions) this isn't clear until your examine Table 12C in detail. (This is further confused because the authors retain the use of the original regions in Tables 7 and 8.)

The values in Table 12C also are confusing because they don't seem to correspond with the need to allocate a reasonable sample to the 5 habitats the 9 regions (3,4,8,12-17) which have a habitat strata. (Table 8 indicates there is no allocation for unvegetated habitat so there may only be 4 habitats.) For example, even with a total of 45 crew years, regions 8 and 15 have only 1 crew year allocated. This implies that there will be a total of 4*2=8 intensive plots using in these regions. Theses 8 plots must be allocated among the 5 habitats. This could result in some habitats only having one intensive plot and there will be no possibility of calculating variances, while for other habitats there will only be 2 intensive plots and the sample size which gives a one degree of freedom estimate of the variance which is clearly inadequate for the use of large sample theory.

I would solve the problem differently. First allocate one crew year per stratum (as was done here). Then try allocating one more sample to each stratum. Choose the allocation which makes the largest CV (across species) as small as possible. This is the marginally best improvement which could be made. Continually add one more sample finding the marginally best improvement to CV for each total crew years. This would produce an allocation within strata which increases monotonically with total allocation.

The problem of sample sizes being too small to calculate variances within habitats may be caused by the sample sizes not being restricted to integers.

Q13: Reviewer #2

It was interesting that the conclusion ended up advocating the base conditions—I don't think that conclusion was totally supported by the data. The greatest relative increase in precision would have come from increasing the number of intensive camps per crew per year from two to three, yet this was judged only a "small" to "moderate" improvement compared to changes in the other parameters. I don't know if an increase in this parameter would be either practical or cost-effective, but it seemed to be somewhat arbitrarily dismissed. Also, I would have liked to have seen the gain in precision that would have resulted from reducing the number of rapid plots per cluster from two to one. An additional step might be to incorporate cost into a subset of the combinations that represent the highest precision. This would provide the most realistic scenario about what was truly the optimal sampling plan.

Q13: Reviewer #3

No response.

Q13: Reviewer #4

The simulation approach is reasonable.

Q13: Reviewer #5

I think so, but I'm not sure that the value for *c* can be so easily separated from this part of the evaluation. Presumably cost trade-offs affect all of these components of the design, and I think it is worth pointing out that increasing *c* seems to have a bigger effect on the quality of the survey than changing any of the other variables. Whether this is actually true, though depends on the relative cost differences between an increase in *c* from, say, 25 to 35 vs. a change in one of the other parameters. In order to fully evaluate what the best scheme would be I think it would be helpful to present the $ cost of each scenario. Then one could assess the relationship between the number of species adequately monitored and the cost of the program. An added advantage of doing this would be that one could make comparisons with the cost of alternative monitoring for certain species. For instance, when there are relatively few species that are added as one moves from Plan B (c = 35) to Plan C (c = 45), it seems very possible that the cost differential might be greater than what it would cost to develop separate surveys for those species (especially, as the species that are added last will probably be those that are rare and local – i.e., species that PRISM is less well suited for, and for which single species surveys might be better). Investigating the trade-off between the economic cost of the program and the number of species adequately covered (especially in comparison to what could be done with other approaches), seems to me to be one of the most important areas for additional work.

Q13: Reviewer #6

Your analysis is essentially a sensitivity analysis of your estimation and variance components derivations to determine the points of diminishing returns for various field parameters. In my experience your conclusions will not conflict with field logistic limitations.

Q13: SUMMARY OF REVIEWERS' COMMENTS

Five reviewers responded. Two suggested that our approach was effective and suggested no changes. The other reviewers made a number of detailed comments that indicate areas that need to be modified and/or explained more clearly (addressed below).

Q13: RESPONSE TO REVIEWERS

Reviewer 2 was concerned that we did not adequately discuss the benefits of increasing the number of intensive camps per crew (n'). Increasing this number would require substantial additional resources, and we feel that similar gains in precision could be achieved more cost-effectively by altering other components of the sampling plan (such as increasing the number of locations sampled by rapid surveys, at the expense of number of plots per cluster). We have elaborated on this in the final manuscript (Bart and Smith *submitted*). We have also clarified that, in the manuscript, the increases in n' are not offset by decreases in other parameters. This leads to an overestimation of the gains in precision, but is unavoidable because of the extreme variability in the cost of additional camps.

While adding a cost factor into the analysis (as suggested by reviewer #5) would greatly increase its utility, this is not practical for this arctic wide analysis. The cost of operating field camps and helicopters varies dramatically across the North, and year to year.

Since the original manuscript was prepared, we have altered the estimation method to acknowledge that sampling in different strata is often not independent (because >1 habitat is often surveyed in a cluster). This change, which corrects a deficiency in the past work, necessitated re-doing this part of the analysis (Bart et al. *submitted*). In doing so, we have addressed the concerns raised by the reviewers.

Bart, J., V. Johnston, P.A. Smith, A. Manning, J. Rausch, and S. Brown. (Submitted) Methods used in the Arctic PRISM surveys. Chapter 2 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

Bart, J., and P.A. Smith. (Submitted) Design of future surveys. Chapter 13 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

14. NEED FOR MORE NATURAL HISTORY INFORMATION

Given the paucity of life history information available for most high arctic-breeding shorebirds, should additional emphasis be directed towards obtaining this information for these species to provide a better basis for developing appropriate survey protocols before the PRISM monitoring program is implemented?

Q14: Reviewer #1

No response.

Q14: Reviewer #2

I'm not sure which species are in question here. It seems that for most of the species targeted in this protocol sufficient information is available on life history to develop PRISM survey protocols. What is lacking, however, is information on breeding distribution, interannual variability in breeding chronology, predation rates, and frequency of re-nesting.

Q14: Reviewer #3

The intensive plot component of the PRISM plan gives just such an opportunity while still serving its primary purpose to calibrate detection rates on the rapid plots.

Q14: Reviewer #4

As the sample design will have to be a compromise among all the species and the stratification do not seem to be vary dependent on species information, I am not sure more life history information will be helpful for the survey design, although it will be very useful to understand their biology.

Q14: Reviewer #5

Hard to say. My sense is that we know enough, in a general sense, to be able to move ahead. But, there does need to be recognition that PRISM might have to evolve as new life-history information becomes available (and explicit protocol for ensuring that period assessments are made). Also, it should be recognized, that this life-history information could be critical for reversing any negative trends that PRISM detects.

Q14: Reviewer #6

No response.

Q14: SUMMARY OF REVIEWERS' COMMENTS

The general consensus seemed to be that there was no need to gather additional information on life histories for individual shorebird species before implementing arctic PRISM. Several reviewers indicate, however, that doing so might be helpful to better understand the species' biology, which may lead to modifications in PRISM protocols in the future and help to develop strategies for reversing any negative trends detected.

The reviewer's answer to this question does not require any changes to the existing PRISM protocol.

Q14: RESPONSE TO REVIEWERS

We are gathering more and more information about species' life history and environment on the intensive plots and will continue to do so. Also, Tier II – the repeatedly surveyed sites - are intended to answer questions about interannual variability in species assemblages, abundance, and survival, as well as re-nesting rates and predation rates, among other things. It is important that the Tier II sites be initiated, but not critical to the success of these arctic PRISM surveys.

15. SURVEY DATES

If the projected number of rapid survey days cannot be completed because of mechanical problems, weather or other unforeseen reasons, should the survey dates be extended in order to sample every plot even if these surveys are later in the breeding season than normal? Or, should the plots not be surveyed? Will reducing the number of plots sampled affect the goals of the monitoring plan? If so, what would you suggest to remedy this situation?

Q15: Reviewer #1

A survey design should be flexible enough to give some leeway for mechanical problems. A survey which can meet the precision target only if every survey is done perfectly on time without mistake is unreasonable. A failure to obtain data from a sufficient number of sites will mean the goals of the plan won't be met. The sample size is already inadequate to justify the large sample theory and further reduction will seriously compromise the quality of the variance estimates. A decision on whether to extend the field season is a biological question which I don't feel qualified to answer. Experience derived from all the years of field trails should provide some guidance on how much work it is practical to accomplish by a crew in a field season. The number of crews deployed should be enough to meet the sample size requirements with provision for mechanical problems.

Q15: Reviewer #2

Survey dates should definitely not be extended because of weather or logistical delays. Reducing the number of plots will affect the monitoring goals. The best remedy would be to extend the number of years required to complete each set of surveys.

Q15: Reviewer #3

No response.

Q15: Reviewer #4

I would like to see them completed unless there are major changes in density (migration) or detection (nesting stage).

Q15: Reviewer #5

I've largely left this question to the Arctic biologists. Though, I would say that it clearly depends a lot on whether the birds will still be nesting – if many are likely to have left then extending the survey period will just introduce unnecessary variation. (If one wanted to be quantitative, and had the appropriate data, then one could estimate the trade-off between this increased variation and the reduction in sample size. Armed with this information, field workers would even be in a position to make an informed decision on a year to year basis.)

Q15: Reviewer #6

No response.

Q15: SUMMARY OF REVIEWERS' COMMENTS

One reviewer felt that the season should be extended until all plots were completed, provided there were no major changes in density and detectability, while another felt strongly that the season should not be extended. Two suggested that our best judgment should be used to determine whether to extend the season or revisit a site in another year.

Q15: RESPONSE TO REVIEWERS

The current analysis of feasibility is based on results from 1,554 plots, 26 sites and 10 years of surveys. While completing these surveys, we have encountered a wide range of weather delays and logistical difficulties. Many of the authors have extensive field experience; we relied on their knowledge and our past experiences to determine the number of plots that could be reliably surveyed by a crew in a season. We feel confident that the goals we've set for a crew year are attainable under all but the most extreme weather conditions. We acknowledge, however, that the Arctic is a place where "the most extreme weather conditions" will be encountered.

In surveys conducted to date, we have sometimes not been able to visit all the rapid plots we had hoped to visit. The important assumption, in such cases, is that the plots we did visit could be viewed as a random sample from all plots in the region. Accordingly it is standard practice to ensure that the entire region is covered even though some plots in each part of the region might be skipped. Furthermore, we are careful to not "cherry-pick" within strata. Thus, we might decide to skip more plots in poor habitat because poor plots, in general, have fewer birds. However, we would not study the habitat maps of plots in the good stratum and then only survey the ones we thought likely to have the most birds.

16. WHEN SURVEYS SHOULD BE CONDUCTED

Should this plan describe in greater detail the timing of surveys with respect to breeding chronology? Should the assumptions regarding survey timing be specifically stated? Given that shorebird breeding chronology is associated with the onset of suitable weather conditions that can vary regionally within years and temporally among years, should this document address how the Arctic PRISM methodology will adjust to these differences in order to conduct the surveys at the appropriate stage of the breeding cycle? Are there conditions when surveys should be avoided altogether because a sizable proportion of the adult shorebirds may not breed?

Q16: Reviewer #1

No response.

Q16: Reviewer #2

I agree that the plan should describe in greater detail how surveys should be timed relative to breeding chronology and how that timing should be determined each year. It should be noted that the length of the breeding season (and synchrony of nesting) will be affected by the earliness or lateness of the spring conditions. These factors should all be addressed for each species, since chronology will vary greatly among them. For some species there may be conditions under which significant proportions of the population may not breed but it is difficult to determine how to handle this. This is related to the question below about large spatial movements of certain populations to portions of the breeding range outside of North America.

Q16: Reviewer #3

No response.

Q16: Reviewer #4

No response.

Q16: Reviewer #5

Yes I think attention to breeding chronology should be made explicit. My big concern is that a systematic shift in breeding chronology (which is quite plausible, e.g., due to climate change) could mask the true population trend. It might be possible to account for this issue simply by using phenology data collected as part of the PRISM protocol (someone who knows the protocol and the birds better than me would be a better judge of this). E.g., maybe you could correct counts relative to an estimate of the sex-ratio (assuming that one sex might become less obvious when nests have eggs), or something of this ilk.

Regarding the final question, I do not think that surveys should be avoided in certain years. Zero is a number! If many birds are avoiding breeding, then we need to know about it, and about how often it occurs. Including these years will be critical to assessing variance, and in terms of population viability the variance in population size can be very important. (There's a Jeremy Greenwood paper on this general topic in Ibis from about 10 years ago; also variance in population size can have an enormous effect on the effective population size, and thus the maintenance of genetic variation in a population, which might be a concern for species with smaller population sizes, or atypical breeding systems).

Q16: Reviewer #6

This whole issue of phonologically induced variation regionally within years and temporally among years wherein a sizable proportion of the adult shorebirds may not breed is, in my experience, the major problem with a survey of this type. Within the 6-10 year duration of a survey rotation, there will be 'bad' years that can indeed be consecutive. I have seen 32 springs and summers first hand on the ground in the western Arctic and the central Arctic while working on geese and swans and I have worked with Peary caribou and muskoxen surveys in the high Arctic subsequently. These 'bad' or 'late' seasons occur quite frequently and they all unfold differently and to different degrees of duration and intensity and differing effects on shorebird populations. In fact, the gradation from 'good' to 'bad' seasons is subtle. The problem is that a 5-10 year survey will incorporate the effect of this fluctuation into the population estimates which is the sum of the strata results which is a sum over years and the trends over the survey rotation will either be accentuated or reduced. The serious question is then, how can one distinguish between a true population change (or no change) and a change seriously influenced by the varying phenology – they are confounded. You have expended a great deal of effort to develop the variance partitions and the subsequent sampling plan analysis with the goal of detecting temporal population changes, but I think that some effort should be directed towards protecting the population estimates themselves from these confounding effects. The only way that I can see to do this is to 'calibrate' each season for the relative density of shorebirds by placing permanent sampling plots in each region (probably just region to be practical). These certainly could be some selection of the intensive plots and sample them each year across the range – not just the strata you are working on in a particular year.

Q16: SUMMARY OF REVIEWERS' RESPONSES

Three reviewers had no comments on this issue. The others recommended that greater detail regarding timing of surveys should be incorporated into the manuscript. One reviewer was concerned that widespread shifts in breeding phenology (e.g. from climate change) could mask the true nature of population changes. He recommended that phenology data should be collected as part of the PRISM protocol to monitor phenological changes over the years. Another felt that extreme weather years could occur for several years in succession and thus obscure our estimates of population change. This reviewer recommended that permanent monitoring plots be established in each region to calibrate weather-related changes in bird populations. Finally, one reviewer strongly recommended retaining non-breeding years in the survey regime, because this information is important for assessing variance in population estimates and in determining effective population size.

Q16: RESPONSE TO REVIEWERS

The issue of the timing of the surveys in an important one. We note however that even in the worst case, inappropriately timed surveys would increase variance but not result in biased estimates provided that the intensive plots are unbiased. Still, the potential for increased variance is significant, and we acknowledge that timing surveys appropriately is sometimes difficult.

PRISM contributors have expended significant research effort to determine the best time to conduct surveys (e.g. Nebel and McCaffery 2003), or when shorebirds might be expected to breed based on local conditions (Smith et al. *In Press*). This latter study examined timing of breeding in relation to local weather and snow cover at four sites over a period of a decade, as suggested by the reviewer. Results suggest that timing of breeding is strongly predicted by snow conditions; a feature that PRISM collaborators have used traditionally to determine the best time to conduct surveys in areas where we have little experience. Further, we continually attempt to improve our understanding of phenology by collecting detailed weather, snow cover, timing of bird breeding, arthropod and plant observations at intensive sites.

In Canada at least, logistical constraints make it difficult to deviate much from a predetermined survey period. So, survey timing must be predicted months ahead of the season, according to "educated guesses" based on average dates of snowmelt in the years preceding the survey year, and average nest initiation dates for the species expected to be present. Final fine tuning of survey timing is sometimes possible using spring reports from researchers in the area, or the advice of intensive crews who may arrive ahead of the rapid crew. Because our flexibility is sometimes limited, our proposed plan attempts to offset potential phenological issues by a) conducting surveys in a given region over two or more seasons; and b) conducting PRISM surveys in the same year at geographically distant locations. Additionally, we advocate the creation of a network of Tier 2 sites that, if implemented, would allow us to study phenology in detail.

Finally, we note that PRISM results to date, and upon which our estimates of power are based, are the product of the surveys that were actually conducted in these arctic regions. Despite any vagaries of weather and timing of breeding, we have successfully carried out surveys at many sites across arctic Alaska and Canada.

Nebel, S., and B. J. McCaffery. 2003. Vocalization activity of breeding shorebirds: documentation of its seasonal decline and applications for breeding bird surveys. Canadian Journal of Zoology 81:1702-1708.

Smith, P.A., H.G. Gilchrist, M.R. Forbes, J.-L. Martin, and K. Allard. In Press. Inter-annual variation in the breeding chronology of Arctic shorebirds: effects of weather, snow melt and predators. Journal of Avian Biology.

17. NEED FOR RECONNAISSANCE SURVEYS

The authors recommend reconnaissance surveys, on which numerous locations are visited briefly to clarify distribution and habitat relationships, as the first step in designing surveys for a given region-habitat strata. Do you agree with this recommendation?

Q17: Reviewer #1

A reconnaissance program could be valuable to eliminate uninhabited areas from the survey program. However, it is important that this process be very conservative and only areas where it is blatantly impossible for any species to nest are removed from the sample frame. Such a process would not have any impact on current calculations of variances which are based on field sites where species area present. However, if a reconnaissance program is used and some portions of the arctic are discarded from sample frame then the extent of these areas should be reported and shown on maps so that users of the data can make an informed evaluation of the survey.

Q17: Reviewer #2

This is a good recommendation, but interannual differences in breeding densities may need to be taken into account for some species.

Q17: Reviewer #3

Yes. It seems a necessary step because of the scale and variation within the areas to be sampled. At least on the Yukon Delta and North Slope of Alaska, variation in shorebird density even within a single CAVM type can be as large as 2 or more orders of magnitude. It may be reasonable to expand this reconnaissance survey into an integral part of the sampling plan, in effect a component even more rapid than rapid-plots.

Q17: Reviewer #4

It seems that reconnaissance surveys are important in defining and mapping the habitat type strata. I expect that the costs will be more than compensated by increased precision.

Q17: Reviewer #5

I lack personal experience on which to assess this question, but it makes perfect sense. If done, however, I would use it as an opportunity to ground truth habitat maps in a formal rigorous way (it sounds as though this is being done, at least in an ad hoc way, already). Note the comments in this section seem to support the idea that habitat maps may greatly over-predict the area of suitable habitat for each species as suggested in my comments on the population size estimates.

Q17: Reviewer #6

Yes, this is vital – nothing in reality is as it seems on paper.

Q17: SUMMARY OF REVIEWERS' RESPONSES

All reviewers endorsed the notion of reconnaissance surveys. One reviewer recommended that it become an integral part of the PRISM protocol, and another recommended that the reconnaissance be formalized to rigorously ground truth habitat maps. One reviewer cautioned that the reconnaissance process should be very conservative; i.e., habitats should not be discarded unless it is impossible for shorebirds to nest in them.

Q17: RESPONSE TO REVIEWERS

Funding proposals for PRISM surveys in Canada formally recognize the need for reconnaissance surveys, and extensive reconnaissance work has recently been carried out in several PRISM regions. The objective of this work was two-fold: to provide preliminary information on habitats, bird densities and bird distributions to assist with the upcoming surveys, but also to develop a formal methodology for conducting this reconnaissance work. Remotely-sensed habitat information plays an important role in population estimation for PRISM, and an important component of the reconnaissance work is to ground truth satellite based habitat classifications. In some instances (e.g. east side of Nettiling Lake, Baffin Island), our reconnaissance work has identified areas virtually devoid of shorebirds, allowing for refined stratification and more efficient sampling of regions. Although some aspects of the methods have yet to be determined, it's clear that reconnaissance work plays an important role in PRISM surveys in areas where little biological information is available.

Details of the methods used to date, and lessons learned from the reconnaissance surveys appear in:

Bart, J., B. Andres, K. Elliott, C. Francis, V. Johnston, R. I. G. Morrison, and J. Rausch. (Submitted) Small-scale and reconnaissance surveys. Chapter 8 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

18. ESTIMATING NUMBERS PRESENT ON INTENSIVE PLOTS

The manuscript describes trade-offs on the intensive plots between using independent observers that will allow for a direct measure of detection rates vs. maximizing the number of plots covered. Do you have advice concerning which option should be followed?

Q18: Reviewer #1

The problem of non-random selection of intensive plots is a practical approach to obtaining an estimate of detection rates. As a consequence the rapid counts on the intensive sites can't be used in the estimation of the average rapid counts. You are forced to make an assumption that the detection rates on intensive count sites are the same as those on random sites. This assumption can't be verified or tested. (See the answer to #12 for an opinion on whether detection rates are a suitable approach to estimation.) The authors state that they want to "obtain detection rates from an adequate number of birds" but really never define what an adequate number is.

If there is a difference in detection rates between experienced and inexperienced observers then it is important that both types of observers be used for the counts on intensive plots. However, this may be difficult since the number of intensive plots is actually very small with a region.

I would probably recommend using approach which maximises the number of plots visited and that a consistent approach be used throughout the range. It is important to consider whether the Lincoln-Peterson estimator will be unbiased given the small number of detections which might be available for many species. There is also the problem of being able to reconcile the observations made by two observers after the field season. It may be easy to do this or it may be extremely difficult. Given that the Lincoln-Pearson estimator is highly sensitive to the reconciliation being done well some field studies should be undertaken before committing to this option.

I think that obtaining detection rates over a larger set of sites is a preferred option than attempting to obtain a 'perfect' count on fewer sites.

Q18: Reviewer #2

In either scenario, it seems that estimates of the total number of birds on an intensive plot are treated as known quantities, without error. If observers work independently, the number of territorial birds detected by both observers is adjusted to reflect the number that was missed, but this uncertainty is not incorporated into the population estimate. If observers work together, the joint total is assumed to be correct; neither a correction factor nor measure of uncertainty is (or can be) incorporated. It seems that both bias and precision should be examined relative to this factor. Both should be evaluated in terms of their effects on power to detect a change in population size.

Q18: Reviewer #3

No response.

Q18: Reviewer #4

On page 24, nonrandom selection of intensive plots is suggest where bird densities is low. I would prefer a random selection from a reduced universe of areas with adequate bird densities and feasible access. Then one could precisely define what part of the population rigorous inference could be made.

Q18: Reviewer #5

I agree with the recommendations that the authors give. Based on their account it does not sound as though the independent-observer approach adds very much, as long as the observers are experienced. Given this, the benefit from the increased sample size seems worthwhile. If situations can be identified where the independent-observer approach does improve counts (e.g., when observers are inexperienced), then it makes sense to use it. I also think that it is sensible to keep a flexible approach to this, albeit with some objective, repeatable way to decide when to use independent observers.

From a pragmatic perspective it would also be good to know how experienced observers best gain that experience – if working with someone else helps them to become good at finding nests more quickly (as I suspect it would), then this might argue further against independent searching. This is because there is the additional trade-off of wanting inexperienced observers to turn into experienced ones quickly. I have no idea if this trade-off can currently be assessed, but it is something that could be considered for future research focused on evaluating the scheme (there might even be information on this in the literature from other species).

[An aside: in section 4.3.1 the question of random site locations is raised. Although the arguments for non-random selection are valid, I think that one would still need to know what proportion of random sites have no birds – this relates to the issue of what is, and is not, suitable habitat. One solution, might be to have a "super-rapid assessment" that can determine for a large number of sites, simply whether birds are present or not. I have no idea how feasible this would be (probably not very!), but somehow this assessment needs to be made (particularly if population size estimates are to be made – if it is not, I think these estimates become especially suspect).]

Q18: Reviewer #6

If your observers are experienced and conscientious and shorebird density is not overwhelming, then more plots would be best.

Q18: SUMMARY OF REVIEWERS' COMMENTS

Two reviewers felt that maximizing the number of intensive plots visited was more important than using an independent observer approach to better determine the number of birds present on the plots. One reviewer felt that the approach should be flexible, and that an independent observer approach may be useful when observers are inexperienced.

Reviewer 1 pointed out that our analysis requires an assumption that detection rates on rapid surveys of intensive plots are equal to detection rates on the truly randomly selected rapid plots. This reviewer also cautioned us that reconciliation must be done well for the Lincoln-Peterson estimates to be unbiased.

Reviewer 3 argued that the precision of our density estimates should be incorporated into the variance of our population estimates.

Q18: RESPONSE TO REVIEWERS

Since the peer review, we have investigated the ability of observers to find nests on the intensive plots in detail (Smith et al. 2009). Our results suggest that the double-sampling requirement of nearly complete counts of birds in the intensive plots is likely to be met for birds with nests that survive over several nest-searching visits. We found no evidence of site effects on the rate at which nests were detected, which may be taken as an indication that non-random selection of intensive plots does not introduce substantial bias.

Based on these results, we now recommend that the potential benefit of an independent-observer approach does not warrant the increased effort in most circumstances. Instead, the focus should be on hiring experienced observers, and obtaining accurate counts from a sample of 4 plots (the number identified as optimal in our analyses) by sharing information. Consequently, reviewers' concerns about the Lincoln-Peterson approach no longer apply. For instance, it would not be possible to incorporate the variance of plots' density estimates into the detection ratio, because there is no means of estimating it when intensive plots are surveyed by non-independent observers.

Reviewer 5 notes that non-random selection of intensive plots, sometimes necessitated by logistical constraints, could influence population estimates by underestimating the quantity of unsuitable habitat. Our large sample of plots surveyed with the rapid method often yield few or no birds, and it is this sample that accounts for unsuitable habitat within the stratum. Provided that the non-random selection of intensive plots does not influence the relationship between rapid and intensive counts, the population estimates remain unbiased. We acknowledge however, that this assumption cannot be adequately tested with the data we have now.

Smith, P.A., Bart, J., Lanctot, R.B., McCaffery, B.J., and Brown, S. (2009) Detection probability of nests and implications for survey design. Condor, 111: 414-423.

19. STRATIFYING INTENSIVE PLOTS

Is it possible that detection probabilities will vary with the densities of shorebirds present on the intensive plots? If so, should these plots be stratified by density categories (for example low/medium/high) to ensure that a representative sample of population densities are found on the intensive plots to correspond with the densities encountered on the rapid surveys?

Q19: Reviewer #1

See the answer to question #12 for comments on the relationship between density and detection rate. The number of intensive sites within a region will be too small to introduce further stratification.

Q19: Reviewer #2

Detection probabilities certainly will vary with densities on intensive plots. I would also expect variance of detection ratios to vary with densities. The interplay between both of these will be important in determining the power to detect change in population size. Using a single detection probability certainly masks variability that should be addressed more directly.

Q19: Reviewer #3

No response.

Q19: Reviewer #4

Detection probabilities will probably vary with a number of factors. It would be prudent to examine the effect of various factors, but it may be limited by sample size.

Q19: Reviewer #5

It is certainly possible …. in fact I think there is evidence from the songbird literature that as singing rates go up, detection rates go down. One could test the idea if one had independent measures of nesting density (as in the study of ours that I described previously). Whether one should stratify depends (a) on whether there is really a problem and (b) on what the trade-offs would be with other things – one can always find new ways to stratify but eventually the gain will diminish. For now, I would say that there is no basis for this kind of stratification, but I would try to address (a) at least to determine if there is even a problem.

Q19: Reviewer #6

If this does present a problem, could you post-stratify?

Q19: SUMMARY OF REVIEWERS' COMMENTS

Reviewers agreed that detection rates probably do vary with density though none said specifically that we should estimate rates separately for different density classes, and some of the reviewers recognized that this would not be feasible.

Q19: RESPONSE TO REVIEWERS

We agree that detection rates could vary by density but this will only cause bias if our sample of plots is non-representative of plots in general. This is a possible danger, and we attempt to select intensive plots in different habitats with variable densities (although a minimum number of birds must be present so as to be able to calculate detection ratios), in order to make them representative.

20. NEED FOR INTENSIVE PLOTS

Given the PRISM monitoring standard (see ms. page 2), do you believe the logistical and financial costs are justified to minimize bias by putting the proposed effort into determining detection probabilities? Could we be just as comfortable using some index method?

Q20: Reviewer #1

It is difficult to answer this without consideration about how the data will be used. Some uses could require an estimate of the population size. Failing to provide it could severely limit the usefulness of the survey. However, it would be useful to examine the precision of the index and quantify the cost of the trade off.

Q20: Reviewer #2

I think the results of these preliminary studies demonstrate clearly that an index of population size would not be reliable for monitoring trends. Not only are the numbers of territorial males underestimated consistently for most species, but also variability among observers in detection ratios is high and must be accounted for.

Q20: Reviewer #3

The PRISM design with both rapid and intensive plots provides a creative and efficient procedure to obtain the bridge between both worlds. The effort that goes into searching the intensive plots, and getting the rapid plot crews to survey them at multiple times, are absolutely necessary

Q20: Reviewer #4

The purpose of monitoring shorebirds is to recommend some action if there are critical population declines. As actions are likely to be expensive, they will likely be resisted and rejected if the evidence is not credible. Failing to account for detectability will give doubters a good reason for not taking any action until the problem is studied further.

Q20: Reviewer #5

Hard to say, because I don't know what the cost trade-offs are likely to be. But, as I've indicated before, I'm not convinced the population size estimates are necessary to do an OK monitoring job (after all, we're all pretty happy, mostly, with the BBS for songbird monitoring and that doesn't produce population estimates). And, as indicated before, I'm not at all sure how accurate the population estimates are likely to be (especially after adjustments to accord with Morrison). It may well be, however, that all the concerns I've raised can be addressed by the authors, as I am sure that they have already considered most of the points I've raised.

Q20: Reviewer #6

The answer to this comes from really understanding why you are doing this project, what the purpose is, and, most importantly, just how much work the numbers that you get in the end are supposed to do for you. The question boils down to just how much is someone willing and prepared to pay for this and whether they can be convinced that the proposed estimates are indeed necessary for decisions that may need to be made – i.e., do we really need population estimates???, are they really overkill? From my point of view, I have no concerns with the design of this project and the mathematical derivations leading to it – they are correct, but they are theory and what matters is what happens in the field. The concern that I do have is, as I have stated above, the uncontrollable noise introduced by the multi-year aspect of each part of the survey, and hence, I don't know if you really will be able to quantify shorebird population sizes across such a massive area and detect changes in population.

My answer to your question is the latter (especially if the money mangers balk) – establishing, right for the get-go, a suite of large permanent plots across the strata, probably in key areas, that are sampled each year *ad infinausium*. Perhaps some sort of analysis of climate-weather records by the weather office folks could give you a good idea of the scale and the temporal and spatial distribution of phonologically 'late', 'normal', and 'good' seasons. This information in combination with shorebird distributions would give you guidance as to where and how many plots you should have. Annual sampling of these plots would give you much more information on shorebirds, their predators, their success, their response to spring conditions, and a direct annual measurement of trend. This approach would certainly be marketable – remember the huge super-collider that the physicists were building in Texas and Congress pulled the plug.

Q20: SUMMARY OF REVIEWERS' COMMENTS

Four reviewers felt that estimates of detection rates were necessary, and that the costs associated with generating them were justified. They acknowledged that population estimates are important for determining when to take action, and argued that an index method would be less defensible when politicians are approached for conservation action. One reviewer was uncertain whether an index might suffice but added that he felt the authors had probably considered these issues in detail. One reviewer recommended permanent plots, surveyed each year, but did not say whether he thought detection rates should be estimated on them.

Q20: RESPONSE TO REVIEWERS

Recent literature on monitoring has emphasized the need to estimate detection rates whenever feasible. In the Arctic, this seems especially important to us given the large changes in phenology, habitat, and observer skill that are unavoidably a part of the PRISM surveys. Double sampling attempts to correct for these sources of bias, and though it is not free from assumptions, we doubt that population or trend estimates lacking estimates of detection rate would be considered credible. Because four of the reviewers agreed with this point of view, and none clearly disagreed, we anticipate continuing to estimate detection rates.

We agree with reviewer #6 that permanent monitoring sites are a critical component of a shorebird monitoring program. We did not discuss the issue in detail in the reviewed manuscript. However, we strongly advocate the establishment of a network of permanent sites in key areas (see Pirie et al. *submitted*). We view these sites as an important part of the larger PRISM initiative, but separate from the arctic-wide surveys discussed here.

These PRISM "Tier II" sites are discussed in detail in:

Pirie, L., P.A. Smith, and V. Johnston. (Submitted) Tier two surveys. Chapter 11 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

21. VARIATION IN METHODS

Logistical issues require PRISM to use slightly different survey protocols in different parts of the Arctic. For example, Canada has two people work together to survey the rapid plots whereas Americans use one person. Some biologists believe regional differences in shorebird densities should allow them to adjust the search effort on intensive plots or perhaps increase the sizes of these plots. Additionally, some species may appear on the rapid surveys but not on the intensive plots within a region, while other species may be so rare that a species-specific detection value cannot be estimated. While equation #3 indicates the need for an overall detection rate but not any condition-specific detection rates, there may be instances when the overall detection rate is based on data from only some of the protocols used to survey a species or on data from another species. In these latter cases, the relationship between the overall detection rate used in the analyses and the actual detection rate for a species in the field may be unknown. Can this issue be ignored, or does it need to be addressed within the PRISM monitoring plan?

Q21: Reviewer #1

If detection rates are fairly large then small differences in detection rates will be unimportant. E.g. using 0.9 instead of 0.95 isn't going to be too influential on the population estimate. If detection rates are small then it is important. E.g. using 0.1 instead of 0.5 will change the population estimate by a factor of 2. From table 10 the average detection rate is about 0.8 which is reasonably high. However, it would take a careful assessment of the magnitude of the local variations from the standard protocol before one could attempt to assess the potential for bias form this source.

Q21: Reviewer #2

Given that the detection ratios appear to be quite variable among species (and certainly are, based on personal experience), this issue should definitely not be ignored. Assuming a single detection ratio for all species will certainly lead to bias (some positive, others negative) for several species. Applying the small SE from a single detection ratio to all species would also likely lead to erroneous conclusions about the power to detect a population change for that species.

Q21: Reviewer #3

No response.

Q21: Reviewer #4

This is a question of how much effort should be dedicated to estimating detectability relative to the effort estimating gross density. Make your best decision based on the simulation study, realizing that there may be problems for some species.

Q21: Reviewer #5

Well, ideally it would be addressed …. but one can always come up with new things that conceivably could generate biases. Without data to determine how big the biases would be I am not sure how this concern could be addressed. Ideally one would do experiments to see if the protocol differences affect the estimates obtained – but I don't know how feasible this would be.

Equally, one would prefer not to use generic detection rates for species that are rare (and thus, perhaps by definition, potentially harder to detect). But, as I've indicated before for species that are so rare that this is a concern, I think PRISM is unlikely to be a good choice of monitoring program anyway. The inability to monitor the rarest species, however, does not detract from PRISM's overall usefulness (again, I would draw the BBS – condor analogy).

Q21: Reviewer #6

No response.

Q21: SUMMARY OF REVIEWERS' COMMENTS

Reviewers agreed that detection rates may vary with protocol but recognized that this is only one of many factors that affect detection rates and that our ability to address them all is very limited. It was also pointed out that with high detection rates, a given absolute degree of variation is less important. One reviewer was concerned that an overall detection rate could lead to an underestimation of the variance in population estimates.

Q21: RESPONSE TO REVIEWERS

It is important that detection rates be estimated whenever data are collected. As long as this is done, detection rate data will be available to compare protocols (though not necessarily for every species). We expect the differences in detection rate due to protocol per se to be minor in comparison to the other sources of variance (but see below). While survey methods may differ slightly across the study area, it is important to remember that we strive for high rates of detection on all rapid plots, and have made every effort to design surveys that achieve an acceptably high rate of detection while still being efficient and safe for workers.

It is also important to standardize how birds are counted, and particularly for those species with polygamous mating strategies. In early years in some locations, the surveyors estimated whether birds had territories centered within the survey plot, while in other locations, birds observed within the plot were simply tallied. These different counting methods resulted in substantially different detection rates, and necessitated separate analyses for Alaska and Canada. The current method, to be applied across all areas, does not require the subjective assessment of territory centroids. This change has made surveys more objective and has stabilized detection ratios.

22. SURVEY SAME PLOTS IN THE FUTURE

The proposed design involves initial and repeated sampling efforts over a "several year" period. Given the monitoring standard of "80% power to detect a 50% decline during no more than 20 years using a two-tailed test, a significance level of 0.15, and acknowledging effects of potential bias", should this plan place more emphasis on sampling the same sites during subsequent surveys to increase the power to detect change over time?

Q22: Reviewer #1

Returning to the same sites in a second survey is an effective technique for improving precision only when there is a strong correlation between counts taken in the two time periods. This will be true for some species and not for others. It depends on the site fidelity of the individual species whether this is useful. It may happen that some species shift location from year to year depending on snow cover or because the site is exhausted.

It would require some repeated measurements done on the same sites over several years before this could be assessed.

Q22: Reviewer #2

I certainly think that it would be a much more powerful approach to replicate the same plots over time. One could even incorporate a rotating panel design, although this might be more complicated than necessary. One would need to obtain estimates of interannual correlation in order to determine the relative increase in power that could be obtained.

Q22: Reviewer #3

The presumed gain in precision with repeated plots and paired t-tests is in practice not realized, at least by my experience. There is enough variation due to change in habitat conditions, weather, movements of birds, or something, so that paired plots are no more similar than unpaired plots in the same areas. Unpaired sampling has the additional advantage of including more independent samples, greater coverage of different areas, and getting better estimates of variance.

Q22: Reviewer #4

I expect that a paired t-test would be more precise than an independent one, because it would remove the site component of variance. Thus there will probably be an advantage to revisiting sites.

Q22: Reviewer #5

Without data on the amount of annual variation within sites and the amount of spatial variation among sites I don't think this can be answered very well. Keeping the same sites from year to year is certainly likely to maximize power, and I would probably lean in that direction. But I could be wrong!

Q22: Reviewer #6

Again, phenology.

Q22: SUMMARY OF REVIEWERS' COMMENTS

Most reviewers suggested that it would be useful to have information on interannual variation within sites before making a decision about whether to sample the same sites during subsequent survey rounds. Four of five reviewers suggested that statistical power was likely to be gained by repeatedly sampling the same sites, although this could vary by species. A fifth reviewer indicated that in his/her experience, statistical power is not gained by sampling the same sites in multiple years. The last reviewer's comment (#6) was hard to interpret.

Q22: RESPONSE TO REVIEWERS

We can see statistical advantages to re-sampling plots or clusters. However, in the Canadian Arctic particularly, there is much to be gained in basic distributional information by going to new areas for each survey. We have not yet repeated plots and it is not likely that we will until after the first round of surveys is completed. However, in the revised power analysis we used BBS data to model the benefit of revisiting plots, rather than selecting a new sample. These results demonstrate the gains in precision that could be expected, and provide some guidance as to whether the increases in precision warrant the decreases in distributional information. This analysis suggested that the target CV_1 is 0.42 if plots are revisited in the second round of surveys, but drops to 0.31 if a new sample of plots is selected. This information is presented in chapter 13 of the monograph.

Bart, J., and P.A. Smith. (Submitted) Design of future surveys. Chapter 13 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

23. *ESTIMATES FOR SPECIES WITH RESTRICTED RANGES*

The manuscript uses regional-habitat stratifications combined with regression analyses to estimate population sizes. However, some species including Hudsonian Godwit, Semipalmated Plover, Ruddy Turnstone, and Sanderling have either patchy distributions or very specific breeding habitat requirements. If these specific habitats cannot be readily delineated using available satellite imagery, or when the breeding distributions of species are poorly known, will the proposed analyses necessarily accurately extrapolate the PRISM density estimates to an overall population estimate? If not, how can the analyses be improved to provide more accurate population estimates for these species?

Q23: Reviewer #1

The survey is theoretically based on a multistage random selection of observation sites. This should ensure that the resulting estimates are unbiased. However, the survey will take several years to conduct. It is conceptual that for species which are highly mobile the same birds could be counted in different sites in different years while other individuals will be missed because they are present on sample sites in years when work is being done elsewhere. These two concepts will balance out provided that decision on where to work in a given year is assigned randomly. The resulting estimates are theoretically unbiased but the variance may be underestimated.

Q23: Reviewer #2

This is a major short-coming of using population size as the parameter to monitor. We know very little, really, about the distributional patterns of most shorebirds in the arctic. A more profitable approach would be to monitor density across a series of plots that are replicated over time. An overall trend could be obtained by weighting by densities. If the primary goal is monitoring population trends, then the design should be such as to maximize power to detect trends. I don't see estimating or monitoring total population size as a very important goal for most species of shorebirds.

Q23: Reviewer #3

No response.

Q23: Reviewer #4

The estimates should be unbiased, but the precision will be reduced by the above factors. If information is available, it can be included in the sample design. Reconnaissance surveys will be important to find patchy habitats used by these species.

Q23: Reviewer #5

I think not. I'm not sure PRISM can be improved to provide estimates for these species – I think that there are some species that are inherently not suited to broad-based multi-species approaches and that these species have to be monitored separately. How one should do this, will depend on the species. For instance, given what we know about PRISM and ISS, it seems silly to me to try to monitor Hudsonian godwits anywhere other than at the two relatively small areas in South America where almost the entire world population is known to winter. For other species, however, it won't be so simple.

Q23: Reviewer #6

The limitations of technology become apparent at the extremes. See my discussion above on permanent annually sampled plots.

Q23: SUMMARY OF REVIEWERS' COMMENTS

Most reviewers recognized that this is a precision, not bias, issue and that small sample sizes may result in the accuracy target not being achieved. One reviewer urged that we "monitor density across a series of plots that are replicated over time" presumably meaning fewer plots and more visits.

Q23: RESPONSE TO REVIEWERS

For the time being, we are not using habitat regression equations to refine our population estimates. The results do not show any improvement with the inclusion of regressions. At some point in the future, there may be instances where habitat regressions do improve our estimates (for example, with very particular habitat preferences, or in areas where habitats are very patchy).

Although the power analyses contain a number of uncertainties, at present they suggest that the High Arctic species will be feasible to survey. Species with restricted breeding ranges, such as Black Turnstones, Buff-breasted Sandpiper and Hudsonian Godwit, might be difficult to monitor adequately without targeted surveys. The potential benefits of targeted surveys for certain species is explored in:

Skagen, S.K., P.A. Smith, B. Andres, G. Donaldson and S. Brown. (Submitted) Contribution of Arctic PRISM to Monitoring Western Hemispheric Shorebirds. Chapter 16 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

24. INFLUENCE OF LARGE SCALE MOVEMENTS

Several arctic-nesting shorebirds (e.g., phalaropes, Pectoral Sandpiper) are known to exhibit large inter-annual fluctuations in abundance that probably reflect movements of birds from one area to another rather than population change. These species also show little or no site fidelity among years. The proportion of birds involved in these movements and their geographic extent are unknown, but some movements to Siberia, Greenland, or other locations outside of the PRISM sampling area are possible. Does the present analyses adequately account for the additional variance in the population estimates associated with these movements? Can population movements into or away from Alaska and Canada from other regions potentially produce bias in the PRISM population estimates?

Q24: Reviewer #1

The survey is designed provide an unbiased estimate of the population within Canada and Alaska averaged over the years of the survey. If this is an inadequate goal for management purposes then the survey should be redesigned.

Q24: Reviewer #2

The present analysis most likely accounts for such movements in terms of increased variance in estimates of breeding population size. The sampling frame purportedly includes breeding areas in arctic North America, so it really doesn't matter if the birds move out of the area or die. If a large proportion is absent (but alive) during the first round of surveys but present during the second round, one would conclude that there has been a population increase. In the context of the North American breeding population, this would be a true increase. In the real biological context of the population, however, this would not provide a very true measure of population trend. Thus, for those few populations that might have significant movements between North America and other continents, it would be better to have a population-wide monitoring program.

It would be important to determine for these species how large of a bias might exist and how likely it would be for such a bias to occur. This would depend primarily on the temporal and spatial patterns of such movements relative to the temporal and spatial patterns of the surveys. If there is spatial clustering of surveys that coincides temporally with movements into or out of those areas, then the two could be confounded and bias could result. Otherwise, it might be likely that movements of birds into and out of areas would be balanced across the 4–6 years needed to complete a round of surveys. Given that the surveys would be clustered geographically within years (for logistical reasons), it might be wise to conduct some presence/absence surveys randomly across the breeding range to detect broad spatial movements for such species.

Q24: Reviewer #3

No response.

Q24: Reviewer #4

These movement will certainly contribute to the variance, but this effect could be modeled if information about the movements is available. It would not bias PRISM estimates of the density in the surveyed area, because the density would change when the birds moved out of the surveyed area.

Q24: Reviewer #5

Such movements certainly could produce biases, but PRISM has an amazingly large sampling frame already, and I'm not sure much can be done to improve it. Moreover, although huge scale movements out of the sampling area are possible, their occurrence (as far as I know) is entirely speculative. The only real solution would be to enlarge the sampling frame, which doesn't seem realistic, or to accept the limitations. I don't think that the concerns here are any greater than those one would have about annual FWS waterfowl monitoring in the prairies.

Q24: Reviewer #6

No, how could it. This is another form of phenology effect. Can population movements into or away from Alaska and Canada from other regions potentially produce bias in the PRISM population estimates? Yes.

Q24: SUMMARY OF REVIEWERS' COMMENTS

Three reviewers suggested that, for the North American sampling frame, the PRISM estimates of population would not be biased by interannual movements of shorebirds. Two reviewers argued that bias could arise if surveys were clustered in space and time, and if the first and i^{th} round of surveys occur over periods when a different proportion of the global population of a species is in North America.

Most reviewers agreed that these interannual movements were not a major concern, and were unlikely to result in serious bias. Two reviewers suggested that the sampling frame could be extended to areas outside North America for species that are suspected to have significant interannual movements. One reviewer suggested that field studies be conducted to assess the extent of such movements.

Q24: RESPONSE TO REVIEWERS

Like the majority of the reviewers, we feel that interannual movements of shorebirds will result in increased variance rather than bias. Our study areas are randomly located across the Arctic, and the survey period spans several years, so the effects of movements should be captured in the variance.

There is little direct evidence of the degree to which these large scale movements occur, and we know of no studies which have identified patterns in these movements over time. We agree that researchers should investigate these movements further, but do not feel that we are in a position to do so unless additional resources become available. These data may be available in the future from comparatively intensively studied areas such as Alaska. If such data become available, the PRISM methods will be re-evaluated for the species in question in light of the new information.

Though we do not feel that interannual movements create serious bias in our population estimates, we agree with reviewer #3 that there could be benefits to expanding the sampling frame to areas outside of North America.

25. PROBABILITY OF SUCCESS

Overall, how likely do you think the plan will be in achieving the accuracy target for the range of species. Please rate the plan from 1 to 5 with a 1 indicating serious doubts and a 5 indicating high confidence that the plan will work. Please explain your ranking.

Q25: Reviewer #1

I would rate this design with a 2.

i) The sample size is too small to justify the large sample theory used in development of the variance equations.

ii) The derivation appears to leave out the correlation between the habitat variables and the rapid counts in the variance equation.

iii) The derivation is based upon the assumption that the regression coefficients are pre-determined.

iv) The allocation of samples to regions seems to be counter intuitive since the allocation doesn't increase consistently with total survey effort. The allocation also appears to be too small to allow reasonable variance estimation.

Some of these problems (ii and iii) can be corrected by reworking the variance derivation. The other problems (i and iv) could be studied through simulation but there is no guarantee that there is a viable solution.

Q25: Reviewer #2

I have strong reservations that this program will meet the monitoring goals for most of the targeted species, and thus give it a rank of 2, as currently designed. Major problems that need to be addressed include: (a) inappropriateness of monitoring number of territorial males for polygamous and lekking species; (b) use of single detection ratio for all species (and power analysis that makes projections based on its small SE); (c) the many problems resulting from trying to estimate total population size (extremely poor range maps; poor knowledge of distribution within breeding ranges; large spatial movements of segments of populations of some species; assumptions of density ratios and stratification by habitat); and (d) high variance associated with rapid surveys because of difficulty in consistently assigning birds to plots, especially at low densities. The program may have more success if some of these problems are addressed.

Q25 Reviewer #3

I rank it 5. The sampling plan will work better than what has been conservatively estimated here in this proposal and power analysis. As data are collected, the habitat models will improve and better analytical methods will be developed to include multiple layers of remote data.

Q25 Reviewer #4

I have some questions about the sample size determination (#2), but if that sample size is appropriate, I think the plan will work (5).

Q25 Reviewer #5

I'd be inclined to rank it a 3 or a 4 overall. For rare species, only a 1 or 2, and perhaps somewhat less also for those species with unusual movement and mating behaviors (like phalaropes) where basic assumptions begin to fall short. For species that are widespread and relatively abundant I'd say a 4. Perhaps more important than simplistic (and not rigorously obtained!) rankings, though, I doubt that it is possible to come up with an approach that is more rigorous, and thus more likely to succeed, at monitoring a large number of species on their Arctic breeding grounds. It is a very ambitious (perhaps audacious) project, but I am confident that it will be better than what we are doing now. It does not, however, provide a complete answer to monitoring Arctic shorebirds (as I suspect the authors would agree).

Q25 Reviewer #6

This plan will generate population numbers and variances and assess population trend, and the mathematics to do that are fine, but because you do not know what the shorebirds are doing from year to year and why you will not be able to reliably label the numbers as a 'population size estimate'. What they are and what they are composed of is unknown. Subsequent analyses and interpretations will then be in doubt. Internally the plan is good, but how it relates to external factors is not good. I must rank it 1-2, but if the annual variations are not as real as I believe them to be then I would give it a 4-5 (a wide straddle on the old fence!).

Q25: SUMMARY OF REVIEWERS' RESPONSES

Reviewers ranged widely in their response to this question. Two reviewers (#3, #4) had high confidence that the survey design will achieve its goals for the majority of species; a third reviewer indicated the design should work fine for the common species but not the rare species. The other three reviewers expressed serious reservations, had low confidence that the survey design will achieve its goals, and had concerns about one or more parts of the program, as follows:

1. One reviewer was concerned about the effects of normal, interannual variation in bird numbers in a given area. If this variation is consistently high, he has low confidence that PRISM will achieve its goal. If interannual variation is lower, he ranks the sampling plan highly.
2. One reviewer had doubts about several aspects of the program, particularly
    a. The use of territorial males as the sampling unit is inappropriate for lekking and polygamous species
    b. It is misleading to use a single detection ratio for all species- and then to base a power analysis on the standard error of that ratio.
    c. The abundance of problems related to estimating total population size (poor maps, poor knowledge of distribution, questionable habitat density ratio assumptions)
    d. Variance associated with rapid surveys likely to be high due to difficulties correctly assigning birds as 'in' our 'out' of plot.
3. One reviewer had serious doubts about statistical aspects of the sampling program, particularly inadequate sample size, inappropriate derivations, and faulty assumptions regarding predetermination of regression coefficients.

Q25: RESPONSE TO REVIEWERS

The proposed Tier 2 of PRISM consists of a number of non-random sites in good habitat where surveys will be conducted regularly. These sites will be located across the PRISM sampling range and so should give a good indication of interannual variation in numbers at a given location. This effort is just beginning under the auspices of the Arctic Shorebird Demographic Network.

It is also worth remembering that interannual variation is an issue of precision, not bias. Our pilot data cover several years and areas. The only way that we are under-estimating precision would be if such inter-annual variation happened to be much less in our years so far than it is on average. There is no more reason to think this is true than that inter-annual variation has been higher, so that precision will be greater in the future than we estimate. We also feel that the problem of inter-annual movements will prove to be significant for little, if any, of the PRISM sampling frame. It only causes trouble if we count the same population in different years in different places, or fail to count them in different years, due to these movements. But in Canada, for example, we propose a 10-year cycle with 2 camps/year. How likely is it that such movements would really lead to such over-counting or under-counting?

We do not anticipate that polygamous species will cause any particular problems, as territory centroids are no longer estimated. Birds present in the plot at the start of the survey are simply tallied up. The one lekking species that we have within the PRISM sampling frame (BBSA) is a species that we believe will meet the accuracy target, but could potentially benefit from focussed studies (Skagen et al. *submitted*).

A single detection ratio was used for the power analysis because we do not yet have enough data from enough areas to generate species-specific detection ratios. However, in the PRISM monograph, we present some results with species-specific estimates of detection and evaluate in detail the effects that these rates have on our estimates of population size.

A number of Reviewer #3's reservations relate to the current poor state of knowledge of arctic breeding shorebird ranges, distributions, and the distribution and locations of shorebird breeding habitat. The authors feel that this is a general limitation that would apply to any program of this type. The current analysis is based on the best available information. Over the course of the first round of PRISM surveys, the quality and accuracy of habitat information will improve dramatically. We intend to review the program after the first round of surveys, with the new habitat and distribution information, to determine if we are still on track to meet our accuracy targets.

Skagen, S.K., P.A. Smith, B. Andres, G. Donaldson and S. Brown. (Submitted) Contribution of Arctic PRISM to Monitoring Western Hemispheric Shorebirds. Chapter 16 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

26. SUFFICIENT JUSTIFICATION TO ADOPT?

Does this manuscript and the accompanying documents provide sufficient justification to adopt this regional monitoring effort that will provide estimates of population change for most species, or should greater consideration be given towards developing a small number of species-specific surveys directed at the high-priority species but would not attempt to monitor all species?

Q26: Reviewer #1

I doubt that it would be possible to develop regional specific studies which would be more accurate for high priority species. However, it is possible that such surveys could be developed. It would take a substantial investment to time to develop these surveys. If this effort were undertaken the predicted precision of the results could be compared with the omnibus design presented here.

Q26: Reviewer #2

I do not think that there is sufficient justification at this time to adopt this regional monitoring effort for most species. I suggest that each of the targeted species be evaluated individually, and that the committee should reconsider what parameter(s) would best be monitored for each. I do, however, advocate that we not restrict monitoring programs to those species with small populations or known population declines. I would advocate a coordinated, international network of regional monitoring sites on the breeding grounds that would not only incorporate geographic variability in population trends for wide-ranging species but also maximize data-gathering for a large suite of species simultaneously.

Q26: Reviewer #3

Species of greater concern today are often those species for which we have more information. The species of unknown status with poorly monitored populations may actually be the species of greater concern 20 years from now.

Q26: Reviewer #4

A general survey will certainly be a compromise among all the species. It should be examined for each priority species to determine if it is adequate for that species. If shorebird species occur in similar habitats, there should be efficiencies in a joint survey.

Q26: Reviewer #5

I think that some form of PRISM should go ahead, because it is important to monitor all species and the authors have made a compelling case that the program can work better than many (including me) suspected. But, I also think that it would be useful to examine the financial cost of PRISM in its most rigorous form to a range of pared-down versions, and to examine what could be done in terms of single-species surveys for that cost difference. Without someone putting dollars on these things though, it is hard to tell what makes most sense. E.g., if the full rigor of PRISM only costs 5-10% more, then it is probably worthwhile. If it triples the cost, I'd be more inclined to use the extra money elsewhere.

Q26: Reviewer #6

I think that it is clear that regional monitoring is vital for shorebirds but not necessarily for all bird species groups (they should be ascertained on a case by case basis) and not necessarily this population estimate plan. You do not address the reasons why estimates of population sizes etc are necessary as opposed to permanent sample plots vis-à-vis who would be willing to pay.

Q26: SUMMARY OF REVIEWERS' COMMENTS

This question raises two issues, first whether this set of regional surveys should be carried out in order to provide population trends for most arctic-breeding shorebirds, and second whether species specific surveys should be considered instead.

On the issue of conducting the broad-scale surveys, there was one reviewer in support and two with qualified support, including one who thought the design should be analyzed for effectiveness at covering high priority species, which is the intent of the PRISM committee, and one who thought that cost estimates should be completed and compared to other approaches. There were also two reviewers who thought the broad scale surveys were not justified or not fully justified, and one who did not answer directly.

On the issue of using alternative approaches such as individual surveys for high priority species, there was general concern from three reviewers: one reviewer noted that species-specific surveys would be very expensive; one noted that shifts will occur in which species are priorities, so apparently supports surveying all species through the proposed methods; and one noted that joint surveys would likely be more efficient. There was support for alternative approaches from two reviewers, including one who recommended that species-specific surveys be conducted, but also that all species should be monitored, without specifying how. The final reviewer thought that the costs of both approaches should be compared.

Q26: RESPONSE TO REVIEWERS

The goal of our proposed plan is to achieve the stated accuracy target for population estimates of as many species of arctic breeding shorebirds as is feasible. Our analyses were an effort to determine what number of species represents a feasible goal. Species-specific survey programs may have a role for monitoring species of conservation concern, or for example, monitoring species with a restricted breeding range or restricted breeding habitats. However, the goal of species-specific surveys is quite different from that of our plan, and the large number of single species surveys would clearly be an inefficient approach to achieve our stated goal.

Some reviewers advocated a different approach; only monitoring a subset of priority species. As noted by other reviewers, we feel that it would be difficult to prioritize species without a broad survey program. Moreover, we firmly believe that an arctic-wide monitoring program will dramatically improve our knowledge of range and habitat use, allowing us to monitor portions of a species population that may have gone unnoticed before. These indirect benefits cannot be measured, but we expect that they will be great.

We have estimated the cost of the Arctic PRISM protocol suggested here (J. Rausch and P. Smith, unpubl. data), and are happy to provide this and compare it to the cost of other programs that may be proposed.

27. DEMOGRAPHIC RATES

Do you think studies that investigate shorebird vital rates as a measure of a species' likelihood of decreasing or increasing would be more or less effective than conducting monitoring surveys?

Q27: Reviewer #1

A direct estimation of the population through a sample survey provides a very fundamental measure of population status. Vital rates would be far more difficult to measure accurately enough to provide effective assessment of population status. Band recovery rate for mallards have millions of records and are barely enough to run population models and data collection on that scale would be impossible for shorebirds.

Q27: Reviewer #2

As I stated at the onset, I think that for site-faithful species, monitoring demographic rates (particularly adult survival and either age ratios or productivity) would be better than monitoring population size or density. Not only would one be able to detect population declines sooner but one would have critical information about what might be causing the decline. For species that are not site-faithful, monitoring of any kind is particularly problematic.

Q27: Reviewer #3

Definitely not. Estimates of production or survival are generally even more complicated by untestable assumptions, sampling problems, and high variability.

Q27: Reviewer #4

Monitoring vital rates provides important additional information about what rates are affected; however, capture-recapture studies are expensive. Because they are labor intensive, it may not be feasible to randomly locate plots, resulting in substantial sampling bias. Considering the site component of variance, the double sampling approach will probably be much more precise.

Q27: Reviewer #5

I really don't know. Intuitively, I think that estimating vital rates has some big advantages. But I also once discussed this issue with Jon Bart, and he totally convinced me that it was not a good solution to trend detection. Sadly, I forget his entire argument, but I think it was based on the sample sizes one would need in order to achieve adequate precision, and that to do as well as is possible from counting birds one would just need a much longer time series.

More generally, I would say that many of the concerns that are raised about PRISM would not go away if one sampled vital rates (e.g., Is the sampling frame appropriate? Are sample sizes large enough? etc.). The one big advantage of measuring vital rates is that they could help one start to get at mechanisms behind declines – but that advantage is only useful if the declines can be detected as well as they could through standard monitoring approaches.

Q27: Reviewer #6

I think that this approach would a different insight to shorebird populations, but implementing it may be more difficult since the vital rates would vary from place to place, from time to time, as functions of density, and in response to varying seasonal conditions. A big job.

Q27: SUMMARY OF REVIEWERS' COMMENTS

Response to this question ranged from "yes" to "definitely not". Five reviewers felt that measuring vital rates would be more difficult and less effective than estimating population size. Three reviewers noted that similar issues of inter-site and inter-annual variability apply to both measurements of vital rates and estimation of population. Four reviewers argued that vital rates provide important information; information that could potentially be used to manage declines. One reviewer argued that measuring vital rates would be a preferable approach for site faithful species.

Q27: RESPONSE TO REVIEWERS

We agree that measuring vital rates provides important information, and that this information may help to determine the cause of declines. We collect estimates of nest survival at all intensive sites, and feel strongly that permanent sites should be established to monitor vital rates and other aspects of shorebird ecology.

However, we do not feel that measuring vital rates to monitor trends is a preferable alternative to estimating population size. We feel that it would be prohibitively difficult and expensive to monitor vital rates at a sufficient number of sites for representative coverage across the breeding range. As noted by reviewer 3, the problem would be compounded for species with low site fidelity or for species which are difficult to capture and band. It would be exceedingly difficult to monitor vital rates in low-density marginal habitats, though these expansive areas often contain significant portions of the breeding population and are predicted by many to be the first areas to show signs of decline.

We feel that this issue is akin to the debate of broad vs. single species surveys. There is unquestionable value in demographic monitoring, and for certain species under certain circumstances, it might be a cost effective solution to simultaneously monitoring population and understanding the life-history stage contributing to the declines. However, we don't see how demographic monitoring could achieve the goal of adequate monitoring, across the breeding range, for most or all species breeding in the Arctic (see Bart et al. *submitted*). If a plan were devised to assess the feasibility of achieving this goal through demographic monitoring, we would review it with interest. In the meantime, we continue to support the idea of monitoring productivity at our intensive camps, and strongly advocate the creation of a network of permanent research sites where more detailed demographic studies could be conducted. We invite proposals to integrate ongoing demographic monitoring with our proposed plan, and to improve our collection of demographic information at intensive camp locations.

Bart, J., S. Brown, R.I.G. Morrison, and P.A. Smith. (Submitted) Other Methods for Estimating Trends of Arctic Birds. Appendix A in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

28. OTHER COMMENTS

Do you have other comments regarding any aspect of either PRISM document? Please provide those comments below.

Q28: Reviewer #1

COVARIATES

The authors incorporate a set of covariates but we are never told what they are. Just the hint "e.g. distance to coast" on page 7. Before being able to pass judgement on the suitability of these covariates and whether the design is adequate it would be useful to understand a lot more about them. How many are included? How much is the variance reduced by including these covariates?

The derivation of the variance fails to include the correlation between rapid counts and the intensive counts. I would suspect that the relationship of counts to covariates is very species specific. Any concerns one has about using an overall detection rate rather than a species specific rates are echoed in a concern about using covariates.

Finally, the derivation assumes the covariates have pre-determined slopes. But there is no indication of how these pre-determined values have been derived. If the slopes of the covariables are to be estimated from the survey results then the variance equation doesn't reflect this and is inappropriate.

NITPICKING DETAILS

Table 2: The definition of c_h includes the incorrect phrase "h=1,…,H"

Table 10: Footnotes 2 and 3 seem to be interchanged

Q28: Reviewer #2

I commend the team of PRISM researchers for their dedicated and tireless efforts to address a very significant void in conservation: a monitoring program for arctic breeding shorebirds. The logistical difficulties are immense and the team's accomplishments are impressive. I am hopeful that my suggestions serve to stimulate discussion and that the North American community can reach a consensus on how best to monitor and ultimately conserve this important component of our avifauna.

Q28: Reviewer #3

The PRISM plan is a truly ambitious, complicated, and expensive program. This is a necessity if these populations are to be monitored. In particular, I think that increased attention is needed to fully incorporate GIS analysis procedures and the interpretation of remote data. This may be a wise investment before the final design is adopted, but not before field work proceeds. More data collection and statistical analysis will undoubtedly lead to opportunities to further improve the design. After an initial period with a few years of field work, complete data analysis and redesign should be scheduled. Before adoption of a long-term monitoring program, this will allow for improvements in stratification, revisions in data collection on habitat conditions and weather, reconsideration of suggested plot search procedures, and adjustment of sampling intensity to include land management concerns.

Q28: Reviewer #4

No response.

Q28: Reviewer #5

My one remaining comment, which I have not really touched on before, concerns the compounding of errors, and thus uncertainty, through all the parameter estimates and extrapolations that go into generating species population estimates (and thus trends). I don't feel that I fully understand what this might mean for the final estimates (though that could be a shortcoming in my understanding, rather than any fault of the authors), but I'd love to see a Bayesian analysis that fully accounted for the uncertainty.

Q28: Reviewer #6

So, to summarize, there are three approaches mentioned here: 1) Population estimation and trend via double sampling – PRISM; 2) Permanent plots sampled annually across range; 3) Population models of vital rates – matrix models.

Q28: SUMMARY OF REVIEWERS' COMMENTS

Reviewer #1 noted that our description of covariates was brief, he stated that our variance derivation failed to account for correlation between rapid and intensive counts, and he reiterated his concern that we used true, not estimated, regression coefficients. He also pointed out concerns with two of the tables. Reviewer #3 urged that more attention be given to finding good GIS layers. Reviewer #5 commented on how errors could be compounded throughout the program and thus lead to biases in population estimates and trends. The other comments complimented the overall effort or summarized past comments.

Q28: RESPONSE TO REVIEWERS

We have stopped using covariates in the analysis, which responds to reviewer #1's first comment. We do not know what reviewer #1 means by saying we failed to account for the correlations been rapid and intensive counts. That correlation is incorporated into the variance of the estimated detection rate (e.g. expressions 10 and 11). His concern about regression coefficients is addressed in our response to question #5. We have corrected the problems he identified with the tables. We agree with reviewer #3 about the need for better GIS layers; this is one of our highest priorities for future work, and we state the need for this in the protocol (e.g. Bart et al. *submitted*)). We feel that we have taken the necessary steps to avoid bias in our population estimates, and do not feel that our sampling design invites the "compounding of errors".

Bart, J., B. Andres, K. Elliott, C. Francis, V. Johnston, R.I.G. Morrison, E. Pierce, and J. Rausch (Submitted) Small-scale and reconnaissance surveys. Chapter 8 in J. Bart and V. Johnston (eds.). Shorebirds in the North American Arctic: results of ten years of an arctic shorebird monitoring program. Studies in Avian Biology.

APPENDIX 1. ARCTIC PRISM PEER REVIEW PROCESS

The following process has been proposed for conducting the peer review of the Arctic PRISM program during the winter of 2002-2003. This process is composed of these steps:

1) Jon Bart and his co-authors will complete the draft overview document for Arctic PRISM.

2) Jon will send this draft document for a "friendly review" to scientists that he selects. He will also provide this document to Rick Lanctot at this time.

3) After the review comments are received, Jon and his co-authors will revise the overview document and produce the final draft. The final draft will be a complete document that includes all necessary information including formula and various data used in the various mathematical calculations. The tentative deadline for completing the overview Arctic PRISM document is early to mid-January 2003.

4) During this "friendly review" process, both Rick and Jon will independently draft a series of questions or topics that they want the peer reviewers to specifically answer. These questions/topics will be sent to me [*Bruce Peterjohn*]. The review panel will also be assembled at this time.

5) I will develop the review packet, which will consist of:

    - The final overview document

    - All current versions of unpublished manuscripts cited in the overview document

    - The questions/topics developed by Rick and Jon. I will consolidate any questions that deal with overlapping topics, and will have to opportunity to add any additional questions/topics that I would like addressed during the review process.

6) This review packet will be sent to the current PRISM steering committee (and other interested parties) for their input. The committee can also suggest additional questions/topics to be included in the review. This process will be conducted under a very short deadline, probably in the range of 7-10 days.

7) Once input from the PRISM steering committee has been received, the final review packet will be sent to the review panel. This peer review will be conducted via email, with each reviewer providing independent comments/responses on the overview document. The review panel will have approximately 4-6 weeks to complete their review and provide comments. The review panel will be free to comment on all aspects of the Arctic PRISM program.

8) Once the comments are received from the review panel, they will be sent to Jon and his co-authors and they will be provided with an opportunity to prepare any "rebuttal" that they believe to be necessary. The review panel comments will also be provided to Rick at this time.

9) Once Jon and his co-authors have prepared their rebuttal, I will develop a list of topics that will need to be addressed by the Arctic PRISM program based on the comments from the review panel and Jon's rebuttal. If this list of topics is extensive or controversial, then I will enlist several others associated with Arctic PRISM to assist in the development of this list.

10) This list of topics as well as the entire set of review comments and Jon's rebuttal will be submitted to the Arctic PRISM steering committee, who will then decide on the process for developing an implementation plan to address the comments. Both Rick and Jon should be involved in this implementation process, if they so choose.