



TECHNICAL REPORT HCSU-007

**SEABIRD MONITORING ASSESSMENT
FOR
HAWAI'I
AND
THE PACIFIC ISLANDS**

John Citta¹, Michelle H. Reynolds², and Nathaniel E. Seavy¹

¹USGS Hawai'i Cooperative Studies Unit,
Kilauea Field Station, P.O. Box 44, Hawai'i National Park, Hawai'i 96718

²U.S. Geological Survey, Pacific Island Ecosystems Research Center,
Kilauea Field Station, P.O. Box 44, Hawai'i National Park, Hawai'i 96718

Hawai'i Cooperative Studies Unit
University of Hawai'i at Hilo
Pacific Aquaculture and Coastal Resources Center (PACRC)
200 W. Kawili St.
Hilo, HI 96720
(808) 933-0706

October 2007

The opinions expressed in this product are those of the author and do not necessarily represent the opinions of the U.S. Government. Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.



Technical Report HCSU-007

**SEABIRD MONITORING ASSESSMENT FOR HAWAII AND THE
PACIFIC ISLANDS**

John Citta¹, Michelle H. Reynolds², and Nathaniel E. Seavy¹

¹ USGS Hawai'i Cooperative Studies Unit, Kilauea Field Station, P.O. Box 44, Hawai'i National Park, Hawai'i 96718

² U.S. Geological Survey, Pacific Island Ecosystems Research Center, Kilauea Field Station, P.O. Box 44, Hawai'i National Park, Hawai'i 96718

CITATION

Citta, J., M. H. Reynolds, and N. Seavy. 2007. Seabird Monitoring Assessment for Hawai'i and the Pacific Islands. Hawai'i Cooperative Studies Unit Technical Report. HSCU-007. University of Hawai'i at Hilo, 122 pp.

Key words: Statistical power, indicator species, monitoring standards, sampling design, Northwestern Hawaiian Islands, time series of breeding seabird abundance

Hawai'i Cooperative Studies Unit
University of Hawai'i at Hilo
Pacific Aquaculture and Coastal Resources Center (PACRC)
200 W. Kawili St.
Hilo, HI 96720
(808)933-0706

October 2007

ACKNOWLEDGEMENTS

We thank D. Dearborn, M. Naughton, E. Schreiber, R. Camp, and K. Brinck for commenting on drafts of this manuscript. Drafts of this manuscript were sent to 28 experts on seabird biology and monitoring with a request for review. These experts were: Bruce Peterjohn, Scott Hatch, William Sydeman, Don Dearborn, William Kendall, Paul Doherty, Paul Sievert, Ian Jones, John Piatt, John Sauer, Dan Roby, Elizabeth Ann Schreiber, Darcy Hu, David Duffy, Vernon Byrd, Beth Flint, Angela Anders, John Klavitter, Brenda Zaun, Holly Freifeld, and Nancy Hoffman. Numerous USFWS field biologists and volunteers provided field assistance with data collection and entry. U.S. Fish and Wildlife Service Biologists E. Flint, J. Klavitter, and N. Hoffman, and M. Naughton provided valuable advice and data management. National Oceanic and Atmospheric Administration, C. Rehkemper and D. Howarth provided logistical support. This study was funded in part by the U.S. Geological Survey Wildlife Program and Science Support Program, and the U.S. Fish and Wildlife Service Office of Migratory Birds. The Hawaiian Islands National Wildlife Refuge, and the Midway Atoll National Wildlife Refuge, and the Remote Islands National Wildlife Refuge collected data. Additional support by was provided by the Hawai`i Cooperative Studies Unit (PACRC, University of Hawai`i at Hilo]. Any use of trade, product, or firm names in this publication does not imply endorsement by the U.S. Government.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
BACKGROUND AND JUSTIFICATION	1
SPECIES SELECTED FOR THIS ASSESSMENT.....	2
STATISTICAL POWER TO DETECT TRENDS IN BREEDING POPULATION SIZES AND CHANGES IN REPRODUCTIVE SUCCESS	5
SUMMARY OF CONCLUSIONS.....	5
SECTION I: ASSESSMENT METHODS FOR EXISTING SEABIRD MONITORING PROTOCOLS.....	9
POPULATION DYNAMICS	9
STATISTICAL POWER	11
METHODS FOR CALCULATING STATISTICAL POWER FOR CHANGES IN BREEDING POPULATION SIZE.....	13
METHODS FOR CALCULATING STATISTICAL POWER OF REPRODUCTIVE DATA	21
SECTION II: SPECIES SPECIFIC ASSESSMENTS.....	23
BLACK NODDY (<i>ANOUS MINUTUS</i>)	23
RED-FOOTED BOOBY (<i>SULA SULA</i>)	30
GREAT FRIGATEBIRD (<i>FREGATA MINOR</i>)	36
LESSER FRIGATEBIRD (<i>FREGATA ARIEL</i>).....	40
RED-TAILED TROPICBIRD (<i>PHAETHON RUBRICAUDA</i>)	41
SOOTY TERN (<i>STERNA FUSCATA</i>).....	49
WEDGE-TAILED SHEARWATER (<i>PUFFINUS PACIFICUS</i>)	52
CHRISTMAS SHEARWATER (<i>PUFFINUS NATIVITATIS</i>).....	57
BULWER’S PETREL (<i>BULWERIA BULWERII</i>)	61
BONIN PETREL (<i>PTERODROMA HYPOLEUCA</i>).....	65
TRISTRAM’S STORM-PETREL (<i>OCEANODROMA TRISTRAMI</i>).....	66
BLUE-GRAY NODDY (<i>PROCELSTERNA CERULEA</i>)	68
GRAY-BACKED TERN (<i>STERNA LUNATA</i>).....	70
SECTION III: SAMPLING DESIGNS FOR ESTIMATING BREEDING COUNTS	73
RECOMMENDATIONS FOR ESTIMATING THE NUMBER OF BREEDING SEABIRDS.....	76
SECTION IV: SAMPLING DESIGNS FOR ESTIMATING REPRODUCTIVE SUCCESS	87
PLOT SELECTION ISSUES	87
SAMPLING DESIGN	88
ESTIMATOR FOR REPRODUCTIVE SUCCESS.....	90
SECTION V: GENERAL MONITORING RECOMMENDATIONS	93
SPECIFIC MONITORING PRIORITIES.....	93
GENERAL MONITORING RECOMMENDATIONS	94
RESEARCH PRIORITIES FOR SEABIRD MONITORING	94
LITERATURE CITED.....	101
APPENDIX A. SAS CODE FOR MONTE CARLO POWER SIMULATIONS.....	107
APPENDIX B. COMPARISON OF PROGRAM TRENDS AND MONTE CARLO SIMULATIONS	111
APPENDIX C. OUTLINE OF MIC ALTERNATIVES	113
APPENDIX D. PLOT NUMBERS FOR MONITORING REPRODUCTIVE PARAMETERS	119

EXECUTIVE SUMMARY

BACKGROUND AND JUSTIFICATION

This document is an assessment of the seabird monitoring program for the United States Pacific Islands (USPI) within the Pacific Region of the US Fish and Wildlife Service (USFWS). This area includes American Samoa, the Phoenix Islands, Palmyra Atoll, Johnston Atoll, Wake Atoll, the Mariana Islands, and the Main and Northwestern Hawaiian Islands. This report was requested by the USFWS to assist in the development of a scientifically sound seabird monitoring program for this region. In this report, we review monitoring methods, analyze existing USFWS data to evaluate the statistical power of current monitoring, and propose recommendations for statistically rigorous seabird monitoring protocols.

This assessment is part of a larger project to design and implement a comprehensive monitoring program for seabirds of the USFWS Pacific Region, which includes the temperate seabirds of the California Current Systems, and the tropical and subtropical species of the USPI. Development of this monitoring program was a priority goal identified in the USFWS Seabird Conservation Plan – Pacific Region (USFWS 2005; Goal 7 page 66).

In September 2004, the USFWS convened a workshop of experts in seabird monitoring in Portland, Oregon to discuss goals and objectives for monitoring seabirds of the California Current System. These goals and objectives were further discussed at a smaller meeting in Honolulu, Hawai`i in November 2004. The information in this assessment of seabird monitoring in the Pacific Islands addresses two goals discussed at those meetings. The first goal was to “detect and understand changes in the status and trends of seabird populations in support of conservation strategies in the USPI.” Four objectives were identified for Goal 1: (1) monitor trends of seabird populations, (2) understand causes of population change, (3) determine conservation status of seabird populations, incorporating abundance, distribution, trends, and threats to seabird populations, and (4) collaborate with partners to achieve and advance all objectives. The second goal was to “integrate seabird monitoring into an overall assessment of the health of marine/coastal ecosystems of the USPI.” Two objectives were identified for Goal 2: (1) use seabirds as indicators of ecosystem health (i.e., structure, function, and productivity), and (2) collaborate with partners to integrate seabird monitoring with other monitoring efforts.

These goals include both monitoring objectives (Goal 1, Objective 1 and Goal 2, Objective 1) and research objectives (e.g., Goal 1, Objective 2). In this assessment, we focus on monitoring objectives, and present analytical tools and data collection protocols that can be applied to the assessment of seabird populations and the use of seabirds as monitors of ecosystem health. We focus primarily on (1) the use of population monitoring (i.e., time series data of population size) to detect temporal trends in breeding population size, and (2) the use of plot-based approaches to estimate reproductive success. This report is complementary to an assessment of analytical tools and data collection protocols for demographic monitoring (e.g., estimating survivorship) of Laysan (*Phoebastria immutabilis*) and Black-footed (*P. nigripes*) Albatross in the Hawaiian Islands (M. B. Naughton, pers. comm.).

This report is organized into five sections. Section I contains background information regarding methods for assessing existing monitoring protocols that are common to all species. Section II contains species specific monitoring assessments. This section contains extensive new analyses of USFWS data to evaluate the statistical power and precision of current monitoring methods. Sections III and IV contain revised protocols and Section V contains general recommendations for monitoring and research intended to help meet the stated goals and objectives. Technical details that are only briefly described in the text are included in greater detail in the Appendices.

SPECIES SELECTED FOR THIS ASSESSMENT

Because it was not feasible to address all of the seabirds (ca. 30 species) that breed in the USPI in this assessment, we selected a subset of 13 species. We have chosen species with a wide variety of feeding habitats and nesting behaviors, with the intent that the information we present about monitoring these species will also be useful for designing protocols for other similar species that might be chosen for monitoring. In choosing species to address in this assessment, we considered three criteria: conservation status (identified in Goal 1), utility as indicators of marine ecosystems (identified in Goal 2), and existing data.

With respect to supporting conservation strategies in the USPI, we prioritized species that were either Birds of Conservation Concern or Stewardship Species. The USFWS has a mandated responsibility to monitor Birds of Conservation Concern (USFWS 2002). These are species that, without some type of management action, could potentially become listed under the

Endangered Species Act as threatened or endangered. Stewardship Species are those species and subspecies for which the USPI supports a large proportion of the global population. Species were considered in this category if over 50% of the global population breeds in the USPI.

Seabirds serve as indicators of marine ecosystems because variation in reproductive success is often correlated with prey abundance at large spatial scales (Montevecchi 1993). Thus, information about seabird reproductive success can be used to monitor spatial and temporal changes in prey composition and abundance (Dearborn et al. 2001). With respect to monitoring marine ecosystem health, we considered foraging guilds and geographic distributions. Species were categorized into five foraging guilds using the classification from Harrison et al. (1983): (1) Albatrosses that forage on the surface by sitting on the water surface and seizing prey, (2) Pelecaniformes that forage by plunge diving for fishes and squid, (3) Tuna birds that forage in flocks above ocean predators (e.g., tuna, porpoises, whales, etc.), (4) Nocturnal petrels that forage primarily on organisms that surface at night (e.g., squid and lantern fishes), and (5) Neuston-feeding terns that forage upon small organisms near the water surface. Although there is undoubtedly overlap among these foraging strategies, guilds are a useful tool for describing the diversity of seabird foraging strategies (Harrison et al. 1983). We selected at least one representative from each foraging guild.

Additionally, we considered the geographic distribution of seabird species. We preferentially selected species that occur in large numbers on National Wildlife Refuges managed by the USFWS, and species with relatively broad geographic ranges. As environmental indicators, seabird species with broad geographic distributions are useful because patterns of spatial variation in reproductive success may be used to detect spatial variation in prey abundance. Furthermore, we expected that these species would be of interest to the widest possible audience.

Finally, we focused on species for which existing data on breeding populations and reproductive success would allow us to quantify the statistical power to detect population trends and variation in reproductive success. For species with little existing data, we encourage biologists to consider the utility of monitoring these little known species and/or developing new methods that can be applied to these species.

Based upon these considerations, 13 species were selected for inclusion in this assessment (Table ES.1). Black-footed and Laysan Albatross were selected, but monitoring protocols for these species were developed separately by USGS Patuxent Wildlife Research Center (M.B. Naughton, pers. comm.), thus they are not addressed in this document.

Table ES.1. Species selected for Hawai`i and Pacific Islands Seabird Monitoring Assessment.

Species	Locations Monitored	Years of Monitoring	Status and Justification
Albatross			
Black-footed Albatross	See USGS report		>95% of the global population breed in USPI; addressed by USGS Patuxent Wildlife Research Center
Laysan Albatross	See USGS report		>95% of the global population breed in USPI; addressed by USGS Patuxent Wildlife Research Center
Pelecaniformes			
Red-tailed Tropicbird	Tern Island, Johnston Atoll, Laysan Island, Midway Atoll, Kilauea Point	1979-present	Data-based selection; Entire US population breeds in USPI
Tuna Birds			
Lesser Frigatebird	No locations		Entire US population breeds in USPI; BCC
Great Frigatebird	Tern Island, Johnston Atoll, Laysan Island, Midway Atoll	1982-present	Data-based selection; Entire US population breeds in USPI
Red-footed Booby	Tern Island, Johnston Atoll, Laysan Island	1983-present	Data-based selection
Sooty Tern	No locations	No data	Indicator of ecosystem status
Black Noddy	Tern Island, Johnston Atoll	1980-present	Data-based selection; Entire US population breeds in USPI
Wedge-tailed Shearwater	No locations	No data	Entire US population breeds in USPI; Indicator of ecosystem status
Christmas Shearwater	Tern Island, Midway Atoll, Kure Atoll	1989-2004	Entire US population breeds in USPI; BCC
Nocturnal petrels			
Bulwer's Petrel	Tern Island, Kure Atoll	1989-2004	>50% of global population breeds in USPI
Bonin Petrel	Tern Island, Midway Atoll	No data	>50% of global population breeds in USPI
Tristram's Storm-petrel	Tern Island, Laysan Island	1995-2002 2004-2005	Entire US population breeds in USPI; BCC
Neuston terns			
Blue-gray Noddy	No locations	No data	Entire US population breeds in USPI; BCC
Gray-backed Tern	Tern Island Midway Atoll	1980-present	>50% of global population breeds in USPI

BCC = Bird of Conservation Concern (USFWS 2002)

STATISTICAL POWER TO DETECT TRENDS IN BREEDING POPULATION SIZES AND CHANGES IN REPRODUCTIVE SUCCESS

We used USFWS data to evaluate two statistical aspects of monitoring. First, we used time series of breeding seabird abundance to evaluate the number of years a population must be monitored to detect a 6.7% annual decline (e.g., 50% decline over 10 years) in the breeding population size with 90% power. Second, we used data from reproductive success plots to evaluate the number of plots needed to generate sufficiently precise estimates of reproductive success. The data for these analyses came from Johnston Atoll and the Northwest Hawaiian Islands (primarily Tern Island, Laysan Island, and Midway Atoll).

Additionally, we discuss the methodological challenges for monitoring these species. This discussion includes topics such as the use of the mean incubation count method to monitor breeding population size and the problem of determining nest occupancy for burrow-nesting species. We make monitoring recommendations and present statistical estimators that are designed for situations encountered in the USPI.

SUMMARY OF CONCLUSIONS

Statistical power to detect trends in breeding population sizes

The results of our analyses demonstrate that the number of years required to detect a 6.7% annual decline in population size varies dramatically both among islands (for the same species) and among species. For some species, such as Red-footed Boobies (*Sula sula*) and Red-tailed Tropicbirds (*Phaethon rubricauda*), this magnitude of trend will be detectable in a reasonable amount of time (ca. 10 years). For other species, such as the Black Noddy (*Anous minutus*), it is much less likely that mean incubation count monitoring will provide useful information about long-term trends because the current counts of breeding populations are highly variable from one year to the next. This variability can be attributed in part to the fact that mean incubation counts do not provide a reliable measure of the number of breeding pairs (observation error), and in part to large fluctuations in the actual number of breeding pairs (process variability). If most of this variability results from observation error, the current methods of monitoring some species will be of little utility. We discuss a number of alternative metrics and pilot studies that may provide better estimates of the breeding population size. However, if much of this variability is process variability, then these fluctuations may provide useful information about population dynamics or the correlation of breeding population size with

oceanic conditions. More information about the relative contribution of process variability and observation error is needed.

Statistical power to detect changes in reproductive success

We recommend continuing a plot-based approach to estimating reproductive success. Our analysis of reproductive success data suggests that for the Black Noddy and Red-tailed Tropicbird, relatively precise estimates (95% confidence interval ca. 10%) can be achieved by monitoring between 6-10 plots. We discuss the estimators that are available for these studies and provide spreadsheets formulas for analyzing pilot data. For burrow nesting seabirds, correctly assessing burrow occupancy is a critical component of monitoring. Therefore, we recommend the use of a “burrowscope” camera to evaluate reproductive success.

Monitoring recommendations

We have made specific recommendations relevant to individual species or specific techniques throughout the document. Additionally, we propose the following set of general recommendations that are widely applicable to a variety of decisions facing managers when designing or implementing seabird monitoring programs. We hope that these recommendations will provide a useful framework for modifying existing monitoring programs and designing new ones:

1) We propose that priorities for monitoring, in order of importance, should be breeding population size, reproductive success, and survival. However, we recognize that in some instances (for example, if the breeding population size is harder to estimate than reproductive success) the order of these priorities may be re-ranked. Although survival is often the most important component of population dynamics (Russell 1999), we have ranked it as the final monitoring priority because it is labor intensive, technically challenging, logistically infeasible at some remote islands, and requires a long-term commitment of resources. However, because survival data are so important, we suggest that banding for other purposes (e.g., the mark-recapture ratio estimator to correct mean incubation count methods for asynchrony, see Section III) be conducted so that adult survival can be estimated.

2) We recommend that the assumption and limitations of the mean incubation counts be considered. This technique is extremely useful, but in some situations (e.g., when many pairs

renest or when surveys cannot be conducted over the entire breeding period) it may fail to provide a useful metric of breeding population size. More information is needed on the relative contribution of process variability and observation error to the total variability of these counts.

3) Plot-based studies, instead of island-wide counts, should be considered for monitoring. Because the entire island can be sampled, does not mean that it has to be. We discuss plot-based methods appropriate for estimating the number of breeding pairs and reproductive success.

4) All plot-level data should be archived in a central location and be easily accessible. The Pacific Seabird Monitoring Database is an excellent mechanism for this task, and we encourage the continued support of this project.

Research recommendations

In a number of cases, we demonstrate that the ability to meet monitoring goals is limited. In some cases our ability to meet monitoring goals is limited because breeding colonies are inaccessible, and cannot be feasibly monitored on a regular basis. In other cases, we simply do not know enough about the breeding biology of particular species to predict how successful the application of monitoring methods will be. Finally, for some questions, especially those that use seabirds as indicators of ecosystem conditions, we may lack basic tools for linking seabirds to the ecosystem characteristics of interest. In all of these cases, additional research has the potential to make significant contributions to seabird population monitoring and the use of seabirds as indicators of ecosystem conditions. We recommend four research topics that may have relevant results for monitoring methods. These research topics are (1) breeding biology and demographics of USPI seabirds, (2) telomere length as a tool for measuring population age structure, (3) corticosterone analysis as a tool for evaluating food availability, and (4) fatty acid and stable isotope analysis as tools for quantifying diet composition.

SECTION I: ASSESSMENT METHODS FOR EXISTING SEABIRD MONITORING PROTOCOLS

The use of long-term avian monitoring programs (e.g., the Breeding Bird Survey) to identify population trends is widely recognized as a powerful tool for bird conservation. As a result, there has been a substantial discussion of statistical analysis of trend data (Nichols 1991, Thomas and Martin 1996, Link and Sauer 1997, Link and Sauer 1998). This work has shown that even common birds may require monitoring for at least 10, and often as many as 30, years before statistically significant trends in population size can be detected. Understanding how many years will be required to detect a statistically significant change in population size is an important component of designing monitoring programs that meet biological objectives (Zielinski and Stauffer 1996).

In this section, we discuss issues and procedures that are common to all seabirds. We begin with a brief overview of population dynamics and demographic modeling. We then discuss how counts of breeding seabirds are conducted and analyzed. This discussion includes information on metrics currently used when counting breeding birds and issues related to power analysis for detecting population trends. Finally, we discuss issues related to the collection and statistical analysis of reproductive data.

POPULATION DYNAMICS

Population dynamics describe the changes in the number of animals over time. Total population size is typically abbreviated as N . This population may be further broken down into subgroups that are defined by age class, reproductive status (e.g., breeders and non-breeders), or other biologically relevant criteria. Change in N represents multiple demographic mechanisms, including birth, death, immigration, and emigration. The proportional change in N from one year to the next is abbreviated by λ . When $\lambda > 1$, the population is growing, when $\lambda < 1$, the population is decreasing, and when $\lambda = 1$ the population is not changing.

A basic approach to calculating λ is to conduct counts of the population size through time (Williams et al. 2002, Doherty et al. 2004). The changes in the population size over the monitoring period can be used to estimate the average annual change. This method is relatively simple, and can often be accomplished without marking individuals or measuring reproductive

success. However, this method does not provide any information about why the population size is changing.

An alternative approach to describing population dynamics is to quantify demographic parameters and then use this information to predict N into the future (Burgman et al. 1993). Typically, this approach is conducted using matrix modeling. Matrix models have two components, a population matrix, in which each entry records the number of individuals in a given age class, and a transition matrix that records the transition probabilities of individuals moving from one age class to the next. These transition probabilities are parameters that describe survival (the probability that an individual remains in the population) and fecundity (the probability that an individual produces young). These parameters can be estimated in the field by monitoring nests to measure reproductive success (see below) and by banding birds to measure survival. Then, by providing information about current population sizes, one can use the model to predict the size of N in the future. One advantage of demographic modeling is that it can be used to identify which demographic parameters are the most important to changes in N . Typically, long-lived organisms tend to have low fecundity, and relatively high survival. As a result, changes in adult survival are usually more important to N than are changes in fecundity (Saether and Baake 2000, Doherty et al. 2004).

Because demographic models require a large amount of information on survival and fecundity, the ability to apply them to all but the most intensively studied species is currently limited. Thus, we have chosen to focus on other approaches to monitoring that can be applied more widely, and with existing data. However, because demographic models are a powerful tool, managers should consider monitoring methods that can be applied to demographic modeling in the future. Analytical tools and protocols for estimating survival of banded seabirds (albatrosses) are presented in a complementary report (M. B. Naughton, pers. comm.).

Reproductive data can also be used to monitor seabird populations. Whereas counts of the breeding population size provide information about the population this year, reproductive success can provide an early warning changes in breeding population size that might be expected in the future (e.g., Schreiber 1980). Additionally, reproductive success can be used to measure the effects of climatic shifts (e.g., Polovina et al. 1994) or shifts in oceanic prey communities

(e.g., Kitaysky and Golubova 2000). Metrics such as breeding success are probably sufficient for such modeling efforts. While identification of the individual would be advantageous, it is not critical. For example, average reproductive success likely declines in El Niño years (when the ocean-atmosphere system in the Pacific is disrupted, influencing fish abundance and distribution) regardless of individual reproductive success. If monitoring food resources is of more interest than population trends, then efforts should focus on reproductive success. The critical question is what level of precision is necessary to meet the biological objectives of monitoring.

STATISTICAL POWER

Regardless of the metric one chooses for monitoring, making inferences about changes over time, or extrapolating from a limited number of plots to the entire colony, raises the issue of statistical power. Prospective power analysis is an analytical technique for assessing study design within a statistical hypothesis testing framework. The backbone of conventional hypothesis testing is the evaluation of null hypotheses. Statistical tests provide a tool to estimate the probability, under the null hypothesis, of obtaining data as or more extreme than those that were actually observed. This probability is the P-value. If the P-value is low, the null hypothesis might still be true, and if the P-value is high, the null hypothesis could nevertheless be false. Thus, hypothesis tests are susceptible to two possible errors; Type I error and Type II error.

Type I error is the probability of rejecting the null when the null is true. This probability is traditionally represented by α (alpha), and for most tests the criteria for statistical significance is $\alpha = 0.05$. Alpha represents the frequency of false positives that we consider acceptable. Thus, at the $\alpha = 0.05$ level, if an experiment was repeated 100 times, we would expect, by chance alone, to reject the null hypothesis five times. Type II error represents the probability of failing to reject the null when the null is truly false, and is represented by β (beta). Statistical power, defined as $1-\beta$, is the probability of detecting a difference when one in fact exists. Alpha and beta are determined by the variance of the data, the difference between the effect size, and the sample size. The details of power have been discussed extensively in statistical and biological literature (Cohen 1988, Steidl et al. 1997, Lenth 2001).

In the case of monitoring changes in population size, prospective power analyses use the estimated variation in abundance to calculate the probability that a defined rate of change can be

detected over a defined time (Taylor and Gerrodette 1993). In general, biometricians agree that prospective power analysis is a useful tool for designing monitoring programs (Steidl et al 1997). However, there is considerable disagreement as to how such analyses should be conducted (See controversy in Gerrodette 1987, Gerrodette 1991, Link and Hatfield 1990). Much of this disagreement involves the statistical assumptions used in the analyses. If these assumptions are not met, then the realized power of a monitoring program may be lower than the prospective power that was calculated. Hence, assessments of a sampling design using prospective power analyses should be considered approximations rather than absolutes.

Appropriate estimates of variance

Variance (or error) is a description of the deviation of each data point from the prediction of a model. As variance increases, power decreases, therefore, an appropriate estimate of variance is critical for accurately predicting statistical power. The total variation can be partitioned into process variation and observation error. Within the context of population dynamics, process variation is the true level of variation in a population. This true variation in the number of individuals is driven by factors, such as fluctuations in food supply or weather related mortality. However, biologists can rarely attain a perfect population count. Observation error is the degree of uncertainty between our estimate and the true value. Unlike process variation, observation error does not reflect true variation in the number of individuals. Numerous authors (e.g., Burnham et al. 1987; Link and Nichols 1994; Gould and Nichols 1998; Mills and Lindberg 2002; White et al. 2002) suggest removing observation error from estimates of total error when making inferences regarding variance in population size. Because of how most seabird data are currently collected in the USPI, it is difficult to separate observation error from process variation. In the following analyses, we have not partitioned the variance into process variation and observation error.

An additional complication when estimating the variance in time series of population size is temporal autocorrelation. For regression analysis to yield valid estimates of variance, the error of each observation must be independent. In reference to population data, the deviation of the count from the model at any time must have no relationship to the deviation of the count during the preceding time period. If the variation in the population is primarily process variation, then this situation is unlikely. For example, if a population is small, there will be fewer young

produced and the population is likely to be small the following year. In such a case, the counts are not independent and estimates of variance will be biased. Alternatively, if the variation is primarily observation error, then there is no reason to expect that this error should carry over to the next time period. As a result, temporal autocorrelation may be much weaker.

Temporal autocorrelation is important for prospective power analysis because if positive autocorrelation structure is present and ignored, then estimates of variance are biased low (Little et al. 2002). Low estimates of variation lead to higher statistical power. As a result, a prospective power analysis that assumes no temporal autocorrelation, when it is in fact strong, would overestimate statistical power.

METHODS FOR CALCULATING STATISTICAL POWER FOR CHANGES IN BREEDING POPULATION SIZE

Data collection

Mean incubation counts. For estimating the power to detect changes in the population size of breeding seabirds, we used counts of nesting birds that were collected on Johnston Atoll and the NWHI. Counting the number of breeding pairs of seabirds in this region poses two major challenges: aseasonality and asynchrony. Asynchrony is variation in the time of nest initiation within years: synchronous breeders will initiate nests within a narrow time window; asynchronous breeders initiate nesting across a larger window. This is not an uncommon pattern in the breeding phenology of many seabirds, for instance in Least Terns (*Sterna antillarum*) it has been shown the nesting often occurs in waves, with many after-second-year birds nesting during early waves and second-year birds nesting during later waves (Massey and Atwood 1981). Aseasonality is variation in the time of nest initiation between years. Seasonally nesting species lay eggs at roughly the same time period each year. Aseasonal species may initiate nesting at any time.

The breeding populations of most seabirds described in this report were quantified with “mean incubation counts” (MICs). The MIC method has proved extremely useful in Northwest Hawaiian Islands (NWHI) as a method to deal with the complications of counting unmarked populations that nest asynchronously and aseasonally. The method is as follows: nests are counted at intervals that correspond to the mean incubation period of a species. Because counts are separated by a time interval that corresponds to the species’ incubation period, it is unlikely

that the same nest is counted twice (unless eggs are not viable and birds remain on the nest for longer the normal incubation period). The main advantages of the MIC method are that: (1) birds do not have to be handled or banded, (2) estimates of breeding population size can be generated with a relatively limited number of counts, and (3) if populations are monitored continuously, then aseasonality is not problematic. The main disadvantage is that it may provide an unreliable measure of the number of nesting pairs when individuals nest more than once during the year.

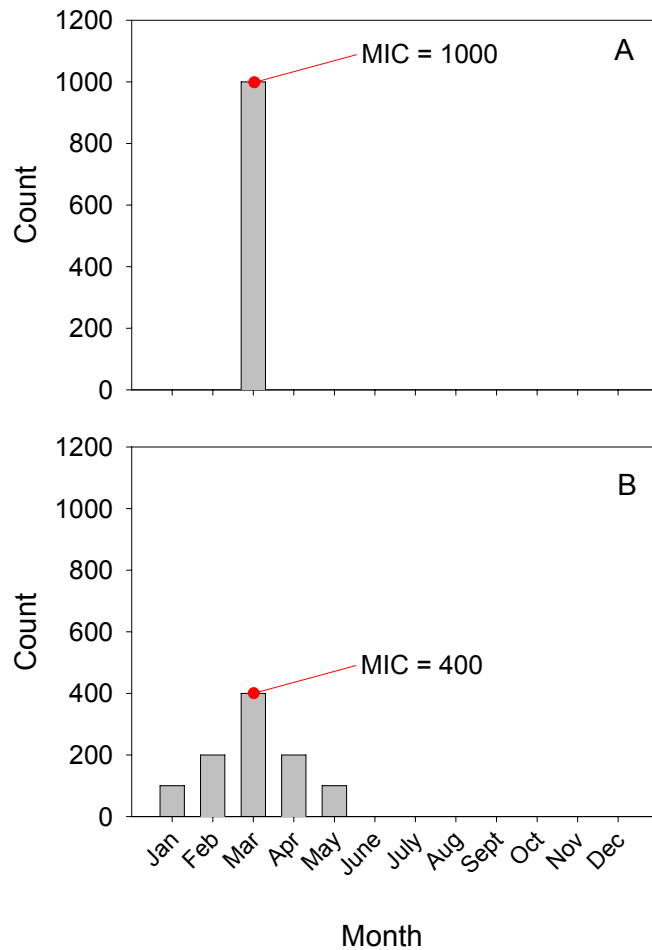


Figure 1.1. If the degree of asynchrony varies from one year to the next, the MIC_{max} will be more variable than the actual population. This pattern is illustrated with a hypothetical situation where the number of breeding pairs does not change, but asynchrony is variable. In panel A, all 1,000 pairs breed synchronously. In panel B, there are still 1,000 pairs, but initiation dates vary. If the interval between counts is 1 month, the estimate of the minimum number of breeding pairs from the MIC_{max} equals 1,000 in case A and only 400 in case B.

If all birds nest only once during the season (no renesting), then the sum of MICs during the season would be an accurate measure of the total number of breeding individuals ($MIC_{total} = \text{all breeding pairs}$), even if the degree of synchrony varies from one year to the next. However, there are a number of problems that may disrupt this relationship (Frederick et al. 2006). If pairs breed more than one once in a single season (hereafter “double-brood”), or renest after failure, then the sum of MICs (MIC_{total}) will overestimate the number breeding individuals ($MIC_{total} > \text{all breeding pairs}$). Alternatively, if nests are initiated, but fail before the next MIC, then the number of breeding individuals may be underestimated ($MIC_{total} < \text{all breeding pairs}$).

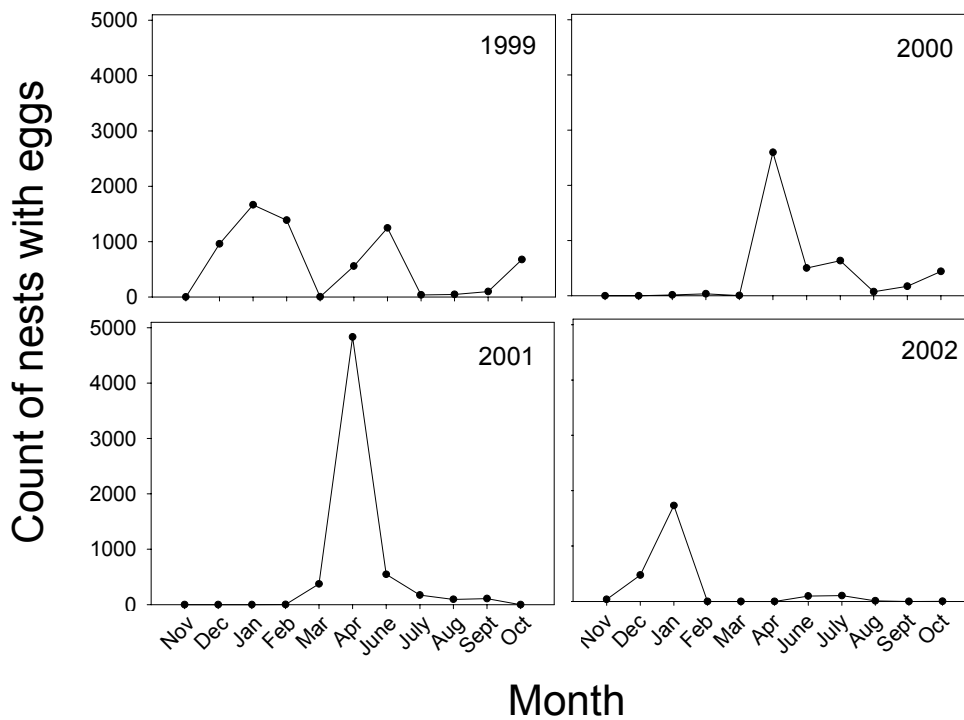


Figure 1.2. Breeding asynchrony of the Black Noddy varies annually. Data are from Tern Island between 1999 and 2002 (this sample is an example – any years could be illustrated). Some years are highly asynchronous (e.g., 1999) and some years are synchronous (e.g., 2002). However, we do not know the degree to which asynchrony can be explained by renesting or double-brooding. For example, it seems unlikely that the true population size more than doubled between 1999 and 2001, as the MIC_{max} would suggest. Thus, this difference between years may be partially, or entirely, explained by a greater proportion of renesting individuals in 2001.

As an alternative metric, one can use the single largest MIC during the season, the maximum MIC (MIC_{max}). This number is a measurement of the minimum number of breeding pairs nesting during that calendar year. Thus, the MIC_{max} provides a conservative estimate of the

number of breeding pairs on an island. However, if nesting asynchrony varies between years, then the MIC_{max} may be much more variable than the population (Fig. 1.1).

As a worst case scenario, consider count data for the most variable species: Black Noddy (Fig. 1.2). Some years are highly asynchronous while others are highly synchronous. The MIC_{max} would be large in years when Black Noddies nest synchronously, but much lower in years when they nest asynchronously, even though there may be little difference in the total number of breeding pairs.

In an earlier draft of this report, reviewers requested a method for quantifying asynchrony. If all pairs nested only once, then asynchrony could be easily quantified as the degree to which MIC values are concentrated (synchronous) or dispersed (asynchronous) throughout the year. Unfortunately, without marked birds, it is probably impossible to distinguish temporal dispersion of MICs that results from truly asynchronous nesting from temporal dispersion that results when pairs either renest or double-brood.

However, we feel that a simple metric may still be of use when illustrating the issue. Hypothetically, we assume that there are 1,000 nesting pairs and that there is no renesting or double-brooding. The MIC_{max} for Figure 1.1B is equal to 400. If we divide the MIC_{total} (1,000) by the MIC_{max} (400), we find that if there is no renesting or double-brooding the population could be 2.5 times (or equivalently 150%) larger than the MIC_{max} (i.e., $1,000/400 = 2.5$), depending upon the year. This number then describes the percentage of all breeders that was captured by the MIC_{max} . When this ratio remains constant from one year to the next (as well as the proportion of birds that renest and double-brood), the MIC_{max} is proportional to the total number of breeding pairs.

To evaluate the variability in this ratio, we computed this statistic for five species based on MIC counts from 1999-2003 on Tern Island. These species were Black Noddies, Red-Footed Boobies, Red-Tailed Tropicbirds, Great Frigatebirds (*Fregata minor*), and Gray-Backed Terns (*Sterna lunata*). The MIC_{total}/MIC_{max} ratios range from 17 to 177% for the Black Noddy, 11 to 68% for Red-footed Booby, 70 to 141% for Red-tailed Tropicbird, 6 to 45% for Great Frigatebird, and 39 to 127 % for Gray-backed Tern.

Again, because we must assume that there is no renesting or double-brooding, this index is not a satisfying metric of asynchrony. Unless individuals are marked, we do not know what proportion of individuals are either renesting or double-brooding. Furthermore, these data represent few years and one island. There is no reason to believe that levels of asynchrony are similar across locations. However, these examples do illustrate that asynchrony may introduce variation in the count data when using the MIC_{max} .

Assumptions of mean incubation counts. We identified two assumptions of MIC methodology. The inferences that can be made using MIC data depend upon the extent to which these assumptions are valid.

1) Data are collected without error. In other words, if counts were repeated, they would be identical. This assumption is probably false, but it is difficult to estimate observation error with a single MIC for each time period. Repeating counts within the same time period could provide an estimate of observation error.

2) The MIC_{max} is consistently proportional to the number of breeding pairs. This assumption is false when birds nest synchronously in some years (MIC_{max} is close to 100% of the total breeding pairs) and asynchronously in others (MIC_{max} is a smaller proportion of total breeding pairs).

These assumptions should be considered when inferences are made using MIC data. MIC_{max} data are probably most representative when species nest synchronously or, if they nest asynchronously, the degree of temporal dispersion is consistent from one year to the next.

Because this section addresses changes in the breeding population size, we have chosen to use the MIC_{max} as our primary metric of the number of breeding individuals. This is the metric that has historically been used by the USFWS (E. Flint, pers. comm.). However, we found that for many populations that have been monitored, using MIC_{max} resulted in statistical power that was below established standards. Therefore, we also investigated the ability of other metrics to increase the statistical power to detect changes in breeding populations.

Monitoring standards for detecting population trends

One of the fundamental questions that must be asked conducting a power analysis is the

magnitude of effect to be detected (Lenth 2001). In the context of population trends, the effect size is usually the annual change in population size. If this change is large, then high power can be achieved with a sample of relatively few years; alternatively, if the effect is small, then more years will be needed to achieve the same power. The difficulty comes in establishing a priori what effect size is important (Cohen 1988, Lenth 2001). Fortunately, both biologically significant effect sizes and the number of years in which they should be detected have been proposed as standards for seabird monitoring.

In developing our prospective power analysis, we considered two standards that have been proposed for Pacific seabird monitoring: (1) the standards proposed by Hatch (2003), and (2) standards proposed by the Alaska Maritime NWR (G. V. Byrd, pers. comm.). Hatch (2003) recommended that monitoring programs be able to detect a 50% decline over 10 years, which is approximately a -6.7% annual decline (on an exponential scale), when $\alpha = 0.05$, with 90% power. The Alaska Maritime National Wildlife Refuge standards are very similar: detect a 50% decline over 10 years (i.e., a -6.7% annual decline on an exponential scale), when $\alpha = 0.1$, with 90% power. This standard differs from that of Hatch (2003) only in that it allows for a higher rate of Type I error ($\alpha = 0.1$ rather than $\alpha = 0.05$). Because power increases as α is increased, fewer years of sampling are required to achieve the same degree of power using the Alaska Maritime NWR standard.

Statistical model

An important component of any power analysis is the selection of an appropriate statistical model. In the section above, we defined the effect as an annual decrease in the population of 6.7% per year. This suggests the simple model,

(Eq. 1.1)

$$N_{t+1} = N_t + N_t r,$$

where N_t is population size at year t and r is the annual rate of change. Unfortunately, often the variance of N increases with the mean of N , violating one of the assumptions of linear regression.

However, this model can be re-written as

(Eq. 1.2)

$$N_{t+1} = N_t e^r,$$

For this model, a natural log transformation of N_t is often used. This offers two advantages. First, the log transformation results in a linear model:

(Eq. 1.3)

$$\log_e(N_{t+1}) = \log_e(N_t) + r,$$

that is easy to fit and interpret. Second, the natural log transformation is often effective in removing the association between mean and variance (e.g., Sokal and Rolf 1995, Thompson et al. 1998, Hatch 2003).

Calculating power

There are a number of widely available programs for power analysis. The most common is TRENDS (Gerrodette 1987, 1991), a user friendly program that allows the specification of either linear or exponential changes in population size. Power is calculated analytically assuming a t-distribution. A major limitation of TRENDS is that only one form of variation can be incorporated into calculations of power (Hatch 2003). The user can account for variation between years or variation within years, but not both. In other words, only observation error or process variation can be explicitly modeled. Hence, complex sampling designs that collect multiple samples within years cannot be assessed. Furthermore, TRENDS only allows the user to account for three functional relationships between variance and population size (See Gerodette, 1993, for descriptions of variance models). TRENDS is currently available on the web at: <http://swfsc.nmfs.noaa.gov/PRD/software/Trends.html>.

Because we wanted the flexibility to allow for any variance structure and serial autocorrelation, we wrote a Monte Carlo program in SAS that allows the user to specify the number of within season counts and the level of variation in those counts (Appendix A). This code can be easily altered to allow for any relationship between variance and population size that is observed and can be expanded to account for serial autocorrelation. To test the base code, we

parameterized the simulations to allow for only one count per season and then compared this directly to program TRENDS. As an example, we used Black Noddy data from Tern Island between 1980 and 2003. The resulting power estimates of these two methods are virtually identical when there is no serial autocorrelation in counts (see Appendix B).

Power analysis of breeding population trends

Assessing error distributions. For each species and monitoring location, we began by assessing if a log transformation was required to standardize the error distribution. To do this, we fitted simple linear and curvilinear regressions and then calculated the residuals for each data point. We then visually inspected the residuals; if the magnitude of the residuals increased with the predicted mean, then we used a natural-log transformation.

Assessing temporal autocorrelation. We used mixed models (Proc MIXED) in SAS and AICc (Burnham and Anderson 1998) to choose between models with different autocorrelation structures. For count data that were equally spaced in time, we examined a first-order autocorrelation structure (AR(1)), where the correlation between adjacent errors extends one time period. We also examined the Toeplitz structure, where all errors separated by a common distance share the same correlation. For count data that were not equally spaced in time, we examined a first-order autocorrelation model with heterogeneity (ARH(1)). This is an extension of the AR(1) model described above, but allows unequal spacing of counts in time. See Little et al. (2002) for a more detailed description of possible autocorrelation structures. We considered autocorrelation structures that were within 2 Δ AICc of the best approximating model.

Calculating power. Using the appropriate Monte Carlo simulation program, we simulated 2500 datasets with the observed pattern of variation (as determined in the above section) and the desired trend (using Eq 1.1 or 1.3 with $r = -0.067$). For each data set, we used a general linear model of the appropriate form (Eq. 1.1 or 1.3) to estimate r and calculate a P-value for the null hypothesis $r = 0$. Power was calculated as the proportion of tests with a statistically significant ($P < 0.05$) negative slope. The number of simulations was chosen as a balance between the consistency of results and computing time. With 2,500 simulations, computing time was modest, yet results were repeatable.

METHODS FOR CALCULATING STATISTICAL POWER OF REPRODUCTIVE DATA

Data collection

For investigating the power to detect changes in reproductive success, we used plot level data collected at Tern Island between 1996 and 2003. Within a season, these plots were visited multiple times, to collect chronological metrics and productivity metrics. Chronological metrics consist of: (1) date of arrival on colony; (2) date first egg observed; (3) date first egg hatches; (4) date first fledged; (5) date last egg observed; (6) date last egg hatched; (7) date last fledged; and (8) date of last departure from colony. Productivity metrics include: (1) hatching success: total chicks per total eggs; (2) fledging success: total chicks fledged per total eggs hatched; and (3) breeding success: total chicks fledged per total eggs. We focused our analyses on breeding success.

Monitoring standards for detecting changes in breeding success

The current monitoring standards are to be able to detect a 20% annual decline (i.e., a 20% change between two years) with 90% power at $\alpha = 0.05$ or 0.1. This standard was originally developed for the Alaska Maritime National Wildlife Refuge based on a subjective evaluation of precision needed for using the data to evaluate correlations with environmental conditions in the marine ecosystem (G.V. Byrd, pers. comm.). This standard was also adopted by the assessment program for the California Current (A. Gall, pers. comm.).

To determine an appropriate standard, we identified three uses of reproductive monitoring data: (1) qualitative investigation of population trends, (2) population modeling, and (3) oceanographic/climate modeling. Reproductive data are likely used in a qualitative fashion (i.e., without rigorous statistics) to assess if populations are stable. If declines in breeding populations are associated with declines in reproductive metrics, then this information can be used to develop working hypotheses to be investigated within more detailed studies. However, a fundamental question must be “how rigorous should our monitoring be for use in developing working hypotheses?” This is a question of precision. Measures of precision indicate the level of confidence in the measurement of a parameter. For example, if breeding success is 0.5 with a standard error (SE) of 0.2, the approximate 95% confidence interval for breeding success is 0.1 to 0.9. This measure of breeding success has low precision. For developing working hypotheses we must ask what level of precision is necessary.

Although we have no standards for the collection of reproductive data, we propose maintaining 95% confidence interval of estimated breeding success of 0.1. With this precision, a difference between year 1 and year 2 that is greater than 0.1 will be statistically significant. We define a 0.1 confidence interval as “desired precision”.

Calculating bias and precision as a function of sample size

We decided to take a reverse approach to assessing the collection of reproductive data. We focus on breeding success (the proportion of all eggs that produce fledged young) because it is the metric of greatest interest to population and oceanographic modeling. We then asked the following questions: (1) Compared to current efforts, how does decreasing effort (i.e., monitoring fewer plots) affect bias in average breeding success? (2) How does precision decrease with decreasing effort? Assessing precision is analogous to assessing statistical power, because the minimum detectable difference between two measurements decreases as precision increases (this is explained more below). Based upon levels of precision and bias, we then assess how breeding success is currently monitored.

We started by calculating yearly mean reproductive success and the 95% confidence interval (using the equations presented in Section IV) for the entire data set of n reproductive plots (here after number of plots is referred to as “sampling effort”). We then randomly selected one plot for removal, and recalculated reproductive success and the 95% confidence interval using this restricted sampling effort of $n-1$ plots. Bias was calculated by using the absolute difference between reproductive success for the restricted sampling effort ($n-1$) and the entire data set (n). We repeated this process as we removed additional plots (e.g., $n-2$, $n-3$), until only two plots were remaining (two plots is the minimum for the estimator). We repeated this process six times for each year, such that our total number of replicates for each sampling effort was six times the number of years in the data set. To summarize these results, we calculated the mean and 95% confidence interval for bias and the confidence interval for each level of sampling effort.

As part of the assessment of sampling methods for seabird reproductive data, we also assessed the basic sampling design and statistical estimators (See Section IV). Section IV also includes a discussion of what plot size is necessary to attain different levels of precision with different levels of variability between plots.

SECTION II: SPECIES SPECIFIC ASSESSMENTS

BLACK NODDY (*ANOUS MINUTUS*)

Data availability

Tern Island. We used 24 years (1980-2003) of MICs for Black Noddies (BLNO) breeding pairs across the entire island (Fig. 2.1). Average incubation length is 30 days, so MICs were conducted approximately once a month.

Johnston Atoll. We used nine years of data, collected between 1985 and 1995 (Fig. 2.1). On Johnston Atoll, it appears that BLNO MICs were collected on plots. It is unclear if a consistent number of plots are sampled each year. As a result, we expected greater observation error in the Johnston Atoll data than in the Tern Island data. For the sake of this analysis, we assumed that data available in the Pacific Seabird Monitoring Database (PSMDB), which includes data summaries until 1995, were collected in a consistent fashion. However, we recognize that such assumptions may not be valid, and emphasize that this is one drawback of conducting power analyses: they will only be as good as the data on which they are based. Since 1995, data collection has been inconsistent, with a smaller number of areas sampled in some years and no data collection occurring in others. Therefore, we excluded years after 1995 from our analysis.

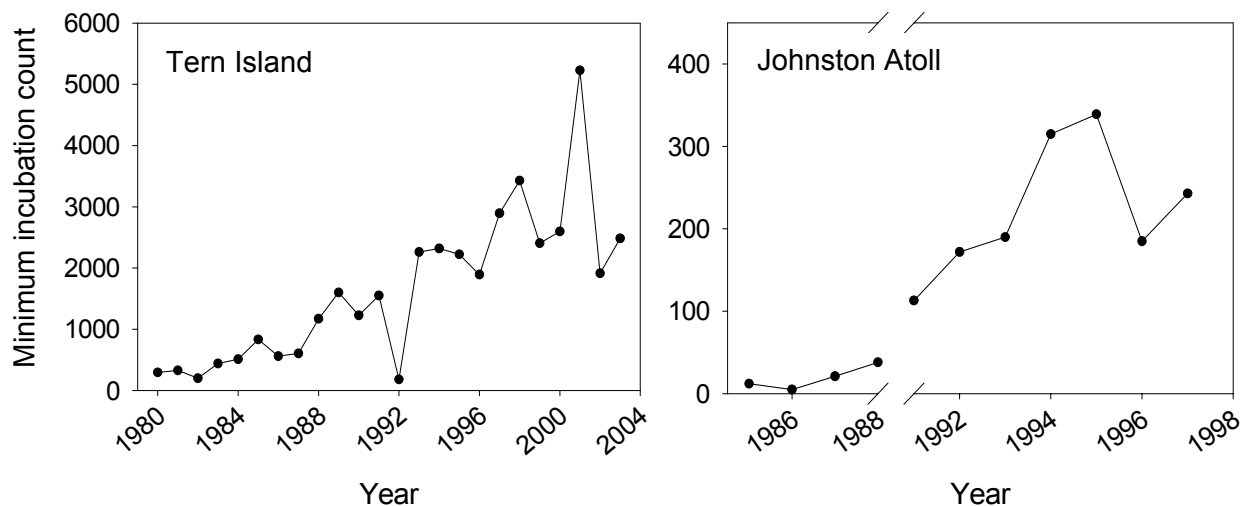


Figure 2.1. Maximum Mean incubation counts (MICs) per year for BLNO on Tern Island (1980-2004) and Johnston Atoll (1985-1998, with 1989 and 1990 missing).

Prospective power of MIC data

Because variance increased with population size, log transformation was used to stabilize errors. Serial autocorrelation was not statistically significant for either Tern Island or Johnston Atoll. For both Tern Island and Johnston Atoll, prospective power was low. With current methods, data would have to be collected for 18 years on Tern Island and 21 years on Johnston Atoll to detect a 50% decline with $\alpha = 0.05$ and 90% power (Table 2.1). Within a 10 year period, only declines greater than 82% for Tern Island and 91% for Johnston Atoll are detectable (Table 2.2).

Table 2.1. Years of observations required to detect trends in Black Noddy with maximum Mean incubation counts (MICs).

Standard	Years of observations	
	Tern Island (SD=0.5539) ^a	Johnston Atoll (SD=0.7073) ^a
6.7% annual decline; $\alpha = 0.05$; power = 0.9 ^b	18	21
6.7% annual decline; $\alpha = 0.1$; power = 0.9 ^c	17	19

^a Standard deviation on a log scale.

^b Recommendations of Hatch (2003).

^c Proposed standards for the Alaska Maritime National Wildlife Refuge (G. V. Byrd, pers. comm.)

Table 2.2. Change in Black Noddy Mean incubation counts (MICs) detectable over 10 years.

Standard	Detectable change over 10 years	
	Tern Island	Johnston Atoll
$\alpha = 0.05$; power = 0.9	82%	91%
$\alpha = 0.1$; power = 0.9	75%	87%

Assessment of current methods

It appears that BLNO populations are highly variable and detection of relevant trends is unlikely (Tables 2.1 and 2.2). This is not unexpected since BLNO are known to re-nest following nest failure and may even double-brood (i.e., fledge two clutches within a breeding season). Using marked birds, Gauger (1999) observed approximately 47% of the population on Tern Island double-brooding in 1987 and 37% in 1988. However, it is unknown how much rates of double-brooding vary over longer time periods (i.e., this is not enough information to conclude that approximately 40% of BLNO double-brood within all breeding seasons).

Breeding is also highly asynchronous and peak breeding times are unpredictable (Dearborn and Anders 1996), hence MICs are expected to be highly variable, even if bird populations are stable. It is unlikely that the true population is as variable as the count of breeding pairs.

Can alternative metrics increase power?

Pre-existing metrics. A number of metrics are collected in addition to MICs. These data include the total number of nests, the total number of nests with eggs, and the total number of nests with chicks. There are a number of reasons to suspect these are better metrics than MICs for indexing the true number of breeding pairs. The total count of nests with eggs may be a less variable index of population size, though the tendency for this species to renest (see above) would suggest that this index is probably an overestimate of true population size. Similarly, if few BLNO fledge two clutches within a breeding season, but all will renest until they produce one successful clutch, then the total count of nests with chicks may be a better correlate of population size and may be less variable.

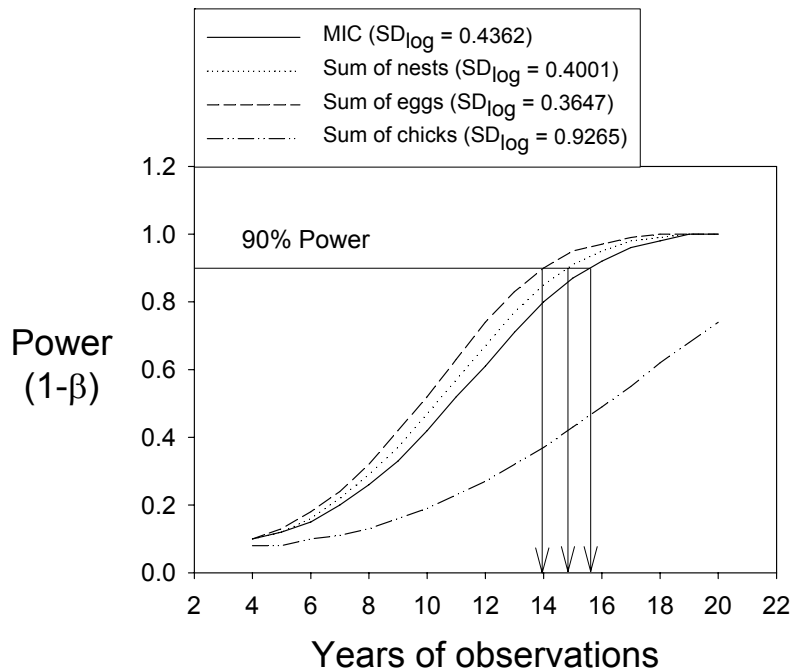


Figure 2.2. Power associated with alternative metrics for counts of breeding pairs for BLNO and a -6.7% annual decline. The standard deviation for each metric, on a log scale, is presented in the figure legend (SD_{log}). Data are from Tern Island between 1999 and 2003. For all metrics, but the sum of chicks, approximately 14-16 years of observations are required to detect trends.

We used Tern Island data (1999-2003) to compare the variance of the MIC with the variance of total nests, total nests with eggs, and total nests with chicks. We then calculated the statistical power associated with levels of variation. For all metrics, except the total number of chicks, approximately 14-16 years of observations are required to detect trends (Fig. 2.3). For the total number of chicks, 20+ years are required to detect trends with confidence. We conclude that these alternative metrics are not a substantial improvement over MIC methods.

Mark-recapture on study plots. Mark-recapture information collected on study plots can be used to determine how the number of total breeders relates to mean incubation counts. This relationship allows us to compute a conversion factor on the reproductive plots and apply this to the entire population. This is accomplished with a *ratio estimator*.

Assume that the total count of breeding pairs is M_{total} , this is the sum of the individual nest counts across a calendar year or breeding season. On each plot, the sum of nests is M_i and the total number of unique pairs, as determined by banding data, is y_i . We divide the number of unique pairs on a plot by the number of total number of nests on a plot to estimate r , the ratio estimate:

(Eq. 2.1)

$$r = \frac{\sum y_i}{\sum M_i}$$

We then multiply the ratio by the total count of nests, assuming that the ratio observed on the plots represents the breeding population on the island. The number of breeding pairs is estimated as:

(Eq. 2.2)

$$\# \text{ pairs} = rM_{total}$$

For example, assume there is only one plot. If observers counted 100 birds on this plot ($M_i = 100$), but there were only 50 unique birds on this plot, the ratio is equal to $50/100 = 0.5$. If the total count is 1,000, then the estimated number of breeders is $r*M$ or $0.5*1,000 = 500$.

A more complete statistical description and a variance estimator are provided under part 6 of Appendix C and in Section III. We have also attached an Excel spreadsheet that illustrates

how the ratio estimator works with example data. This spreadsheet can easily be adapted for use on specific islands.

Although this method is promising for highly variable species such as the Black Noddy, and is expected to increase statistical power, there is no pilot data for testing this method. Hence, we do not know how large the increase in power will be.

Precision of reproductive data

Our simulations suggest that fewer BLNO plots can be sampled with little increase in bias. If only five plots were sampled, bias in our estimate of breeding success would increase by approximately 2% (Fig. 3). The 95% confidence limit for this increase is approximately 7%. Likewise, sampling fewer plots sacrifices little precision. Decreasing the level of plots sampled from 10 to seven results in confidence intervals that are approximately 0.10 at the 95% level (Fig. 4). At five plots, confidence intervals increase to approximately 0.13 at the 95% level.

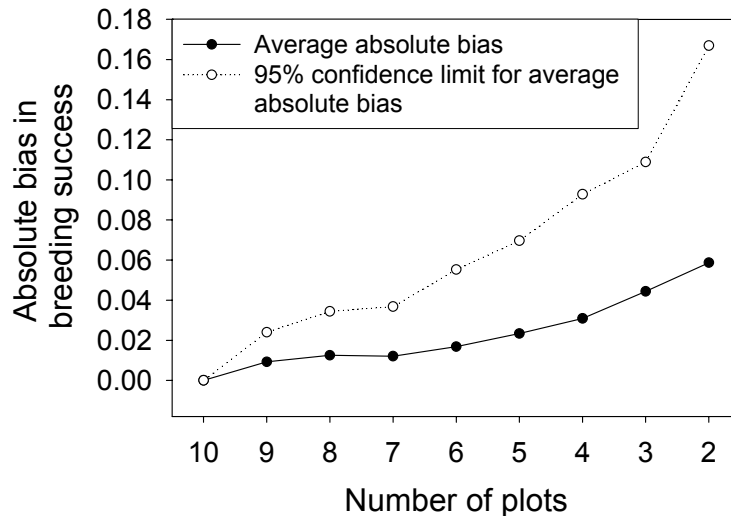


Figure 2.3. Average relative bias and 95% confidence interval versus the number of plots on which Black Noddy breeding success is quantified. All values are in relation to a total of 10 plots. For example, reducing number of plots from 10 to 7 would result in breeding success being biased by 0.01 on average with a 95% confidence limit of 0.04. If breeding success were 50%, then reducing the number of plots would result in an estimate of 46-54%, as compared to 10 plots. The somewhat erratic shape of the curves is caused by having few years of data (1999-2004) and not from the small number of simulations.

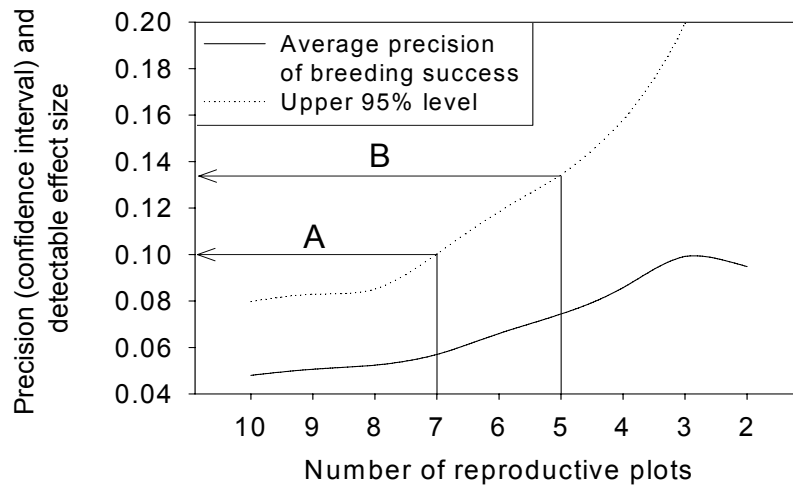


Figure 2.4. The relationship between the number of plots and precision or detectable effect size for breeding success of BLNO on Tern Island. Average precision and the 95% confidence interval are shown. To illustrate the relationship between the number of plots and precision, two drop lines are shown. In case A, decreasing the number of plots from 10 to seven results in a minimum detectable effect size of 0.1 with 95% confidence. In case B, decreasing the number of plots from 10 to five results in a minimum detectable effect size of approximately 0.13. The somewhat erratic shape of the curves is caused by having few years of data (1999-2004) and not a small number of simulations.

Summary and recommendations

Population trend data. Based upon data from Tern Island and Johnston Atoll, we conclude that the ability to detect trends in Black Noddy breeding populations is below desired standards (Tables 2.1 and 2.2). We investigated a couple of alternatives, including (1) using alternative metrics that are pre-existing, and (2) a combination of counting and limited mark-recapture methodology (i.e., the mark-recapture ratio estimator). By using alternative metrics, we were able to reduce the number of years required to detect the desired trend from 18 years to 14-16 years, but we were unable to achieve the goal of detecting the trend within 10 years. The mark-recapture ratio estimator is the most promising method. This method is statistically defensible, will reduce observation error, and will yield information required for population models. However, we do not know how large any gains in statistical power will be, because there are insufficient pilot data to test this method.

With regards to count data, we suggest there are three courses of action: (1) cease counting BLNO, (2) continuing counting BLNO, but with the awareness that 14-18 years will be necessary to detect a 6.7% annual decline, (3) combining MIC methods with limited mark-recapture methods on study plots. If the mark-recapture ratio estimator is used to correct count

data, we suggest that the number of plots sampled be determined by how many plots are required for the ratio estimator to have a CV of approximately 5%. With a population size of 1,000, this would yield an approximate 95% confidence interval of ± 100 . There is no biological rationale for this standard, but this level of variation seems reasonable.

Additionally, if BLNO monitoring is continued, we recommend that detectability of nests be assessed (See Recommendation 1 in Section III). However, detectability is likely a minor issue when compared to the fundamental problems associated with MIC methods for BLNO. We suspect that asynchrony is a larger source of sampling variation in count data than is detectability. Although detectability must be considered, correcting for detectability does not correct for asynchrony.

Reproductive data. Desired precision (95% confidence interval = 0.1) can be accomplished by sampling fewer plots than are currently sampled. Sampling seven plots would satisfy this requirement with 95% confidence. Sampling six plots would satisfy this requirement with approximately 90% confidence. Note that these results are specific to Tern Island. They use the actual variability between plots on Tern Island and the number of nests found within plots on Tern Island. For a more general description of how to estimate breeding success and a discussion plot number/size issues, see Section IV.

We suggest if the variability in the count of breeding pairs can be improved, then reproductive data may be of secondary importance. Attaining reliable counts of breeding pairs should receive priority. If the mark-recapture ratio estimator is used to correct count data, we suggest that the number of plots sampled for reproductive data be determined by the sampling requirements for the mark-recapture ratio estimator. For example, if only five plots are required for the mark-recapture ratio estimator, then we suggest that the precision of reproductive data be compromised in favor of gaining reliable counts of breeding pairs. Even when small numbers of plots are sampled, there is small bias in estimates of breeding success and tolerable levels of precision (Figs. 2.3 and 2.4). If reducing the variability in breeding population counts is not feasible, then we suggest that loss of precision in reproductive success estimates should be minimized, as this metric may be the best indicator for the species.

RED-FOOTED BOOBY (*SULA SULA*)

Data availability

Tern Island. We used 26 years (1979-2004) of MICs for Red-footed Booby (RFBO) breeding pairs across the entire island (Fig. 2.5). Average incubation length is 45 days.

Johnston Atoll. We used 19 years (1983- 2001) of MICs for RFBO breeding pairs across the entire island (Fig. 2.5).

Laysan. We used 13 years (1991-2004) of RFBO data (Fig. 2.5). Data were not collected in 1996. These data were not collected with standard MIC methodology; instead the number of breeding pairs across the entire island was counted once each year.

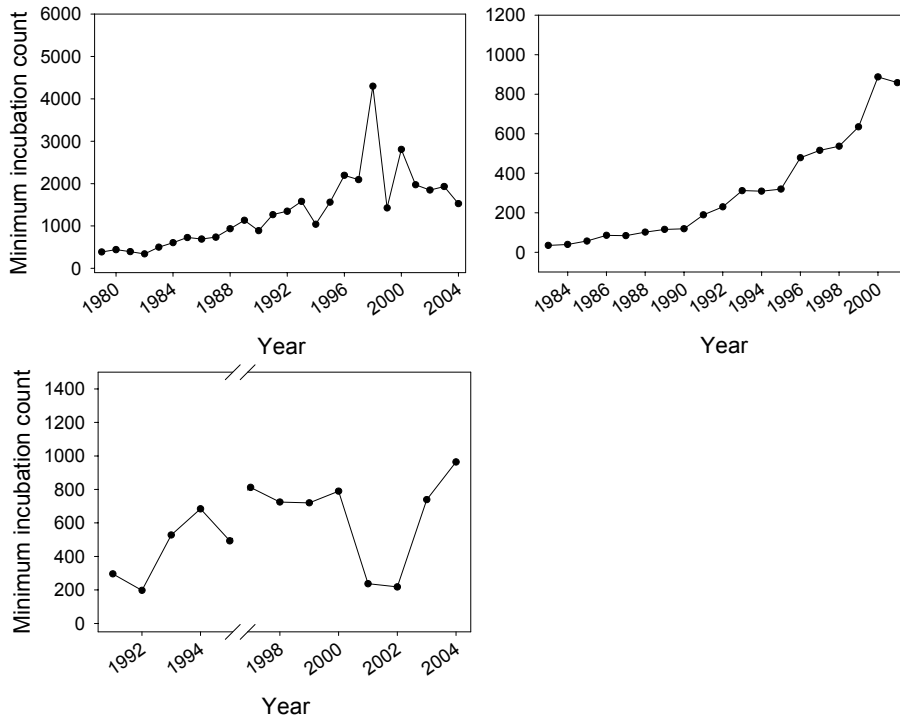


Figure 2.5. Red-footed Booby Maximum Mean incubation counts (MICs) on Tern Island (1979-2004), Johnston Atoll (1983-2001), and Laysan Island (1991-2004, 1996 is missing).

Prospective power of MIC data

Variance increased with population size on Tern Island and Johnston Atoll. Log transformation stabilized errors. On Laysan, normally distributed errors fit the data better than log transformed errors. We could detect no serial autocorrelation in the data series for any

islands. With current methods, data would have to be collected for 18 years on Laysan, 13 years on Tern Island, and 10 years on Johnston Atoll (Table 2.3). Within a 10 year period, declines greater than 82% for Laysan, 63% on Tern Island, and 39% on Johnston are detectable (RFBO Table 2.4).

Table 2.3. Years of observations required to detect trends in Red-footed Booby with Maximum Mean incubation counts (MICs).

Standard	Years of observations		
	Tern Island (SD=0.3055) ^a	Johnston Atoll (SD=0.5598) ^a	Laysan (SD=0.1555)
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.05$; power = 0.9 ^b	13	10	18
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.1$; power = 0.9 ^c	12	9	17

^a Standard deviation on a log scale.

^b Recommendations of Hatch (2003).

^c Proposed standards for the Alaska Maritime National Wildlife Refuge (G.V. Byrd, pers. comm.)

Table 2.4. Change in Red-footed Booby Maximum Mean incubation counts (MICs) detectable over 10 years.

Standard	Detectable change over 10 years		
	Tern Island	Johnston Atoll	Laysan
$\alpha = 0.05$; power = 0.9	63%	39%	82%
$\alpha = 0.1$; power = 0.9	59%	32%	75%

Assessment of current methods

Counts of breeding Red-footed Boobies exhibit different levels of variability among islands. Laysan has the most variable counts, but we suspect this is because the population is only counted once each year. If observers miss the peak laying period, the count will be low. We suspect that levels of variation observed on Tern Island and Johnston Atoll are more representative. Johnston Atoll meets the standards for detecting trends, but Tern and Laysan Island do not.

Can alternative metrics increase power?

Pre-existing metrics. We used data from Tern Island (1999-2003) to compare the variance of the MIC with the variance of total nests, total nests with eggs, and total nests with chicks. We then

calculated the statistical power associated with each metric. For all metrics, approximately 9-11 years of observations are required to detect trends (Fig. 2.6). We conclude that using these alternative metrics only marginally increases power.

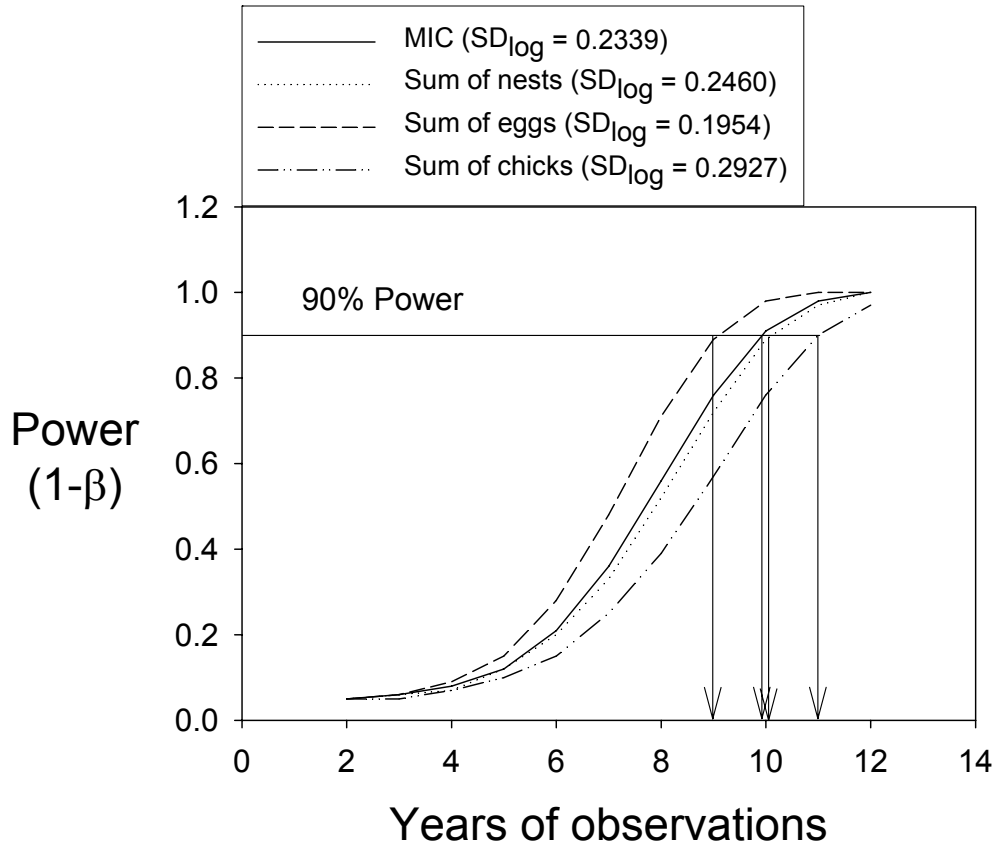


Figure 2.6. Power associated with alternative metrics for population counts for RFBO and a -6.7% annual decline. The standard deviation each metric on a log scale is presented in the figure legend (SD_{log}). Data are from Tern Island between 1999 and 2004. All metrics, except the sum of nests and sum of chick nests, exhibit less variation than the MIC. Using the sum of egg nests decreases the time to detect the trend by 1 year. Note that these results are specific to Tern Island. They use the actual variability between plots on Tern Island and the number of nests found within plots on Tern Island. For a more general description of how to estimate breeding success and a discussion plot number/size issues, see Section IV.

Mark-recapture on study plots. This method is described in detail for Black Noddies and in Appendix C. Because RFBOs are asynchronous breeders (Fig. 2.7), combining mark-recapture with MIC methodology is expected to decrease variance. However, there is no pilot data for testing this method. Hence, we do not know how large the increase in power will be.

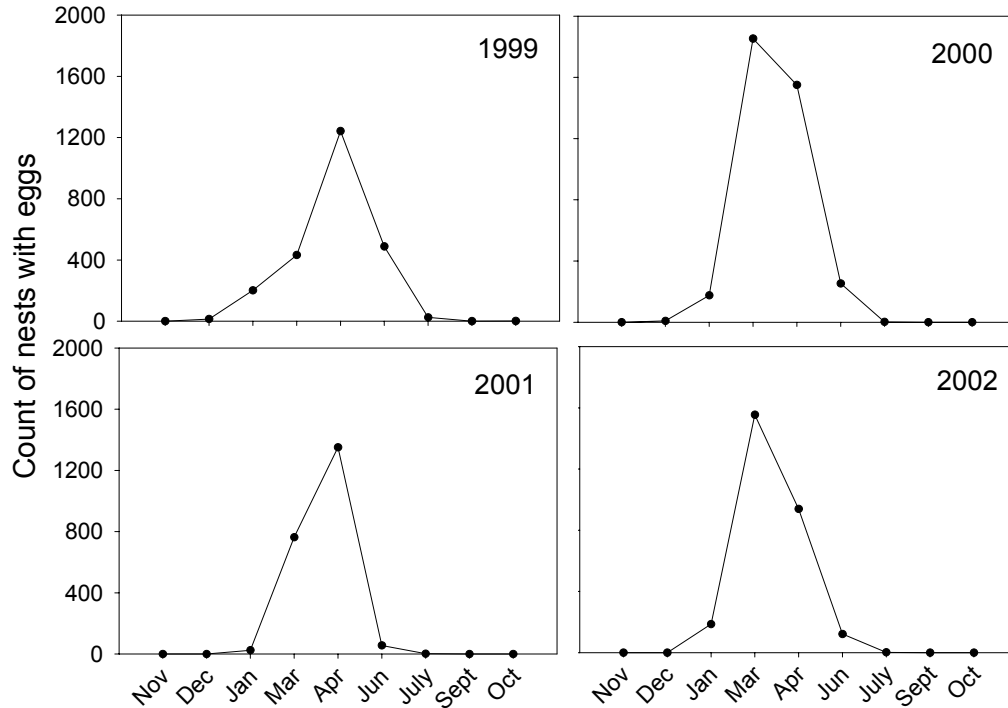


Figure 2.7. Red-footed Booby Maximum Mean incubation counts (MICs) by month on Tern Island between 1999 and 2002.

Precision of reproductive data

Our simulations suggest that fewer plots can be sampled with little increase in bias. If only five plots were sampled, bias would increase by approximately 1.5% (Fig. 2.8). The 95% confidence limit for this increase is approximately 4%. This level of bias is trivial. Likewise, sampling fewer plots sacrifices little precision. Decreasing the level of plots sampled from 10 to five results in confidence intervals that are approximately 0.09 at the 95% level (Fig. 2.9).

Summary and recommendations

Count data. We conclude that the ability to detect trends in RFBO breeding populations is above standard at Johnston Atoll, near standard at Tern Island, and below standard at Laysan. We suggest that standardizing counting methods at Laysan (i.e., use MIC methodology) might improve power to detect trends. Because power is near standard on Tern Island, the mark-recapture ratio estimator is expected to improve power to within standards. If only one

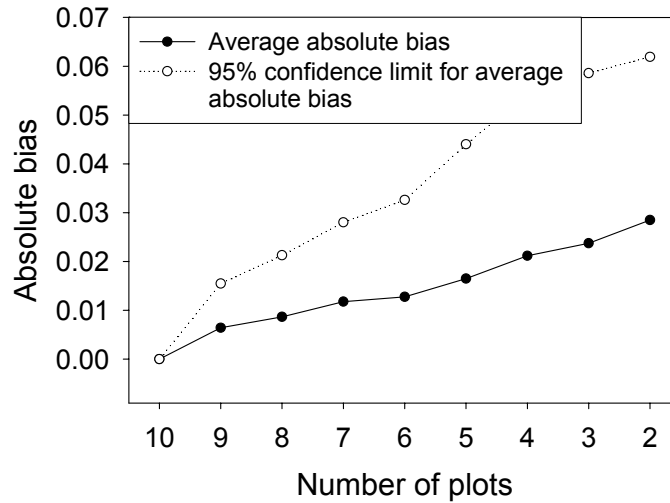


Figure 2.8. Average absolute bias and 95% confidence interval versus the number of plots on which RFBO reproductive data are collected. All values are in relation to a total of 10 plots. For example, reducing number of plots from 10 to five would result in breeding success being biased by 0.015 on average with a 95% confidence limit of 0.045. Note that these results are specific to Tern Island.

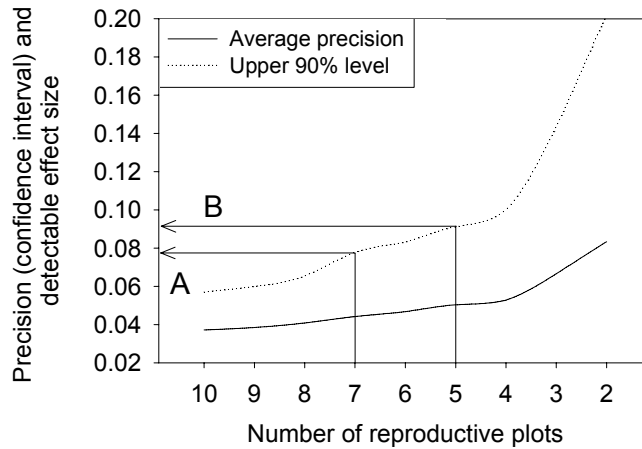


Figure 2.9. The relationship between plot number, precision, and detectable effect size in simulations of RFBO breeding success data from Tern Island. Average, 95%, and 90% confidence intervals are shown. To illustrate the relationship, two drop lines are shown. In case A, decreasing the number of plots from 10 to 7 results in a minimum detectable effect size of 0.08 with 95% confidence. In case B, decreasing the number of plots to five results in a minimum detectable effect size of approximately 0.10. Note that these results are specific to Tern Island.

species is selected as a test for the mark-recapture ratio estimator, we suggest that RFBO are more appropriate than Black Noddies. RFBO are less numerous and will be easier to mark and keep track of. Furthermore their breeding populations are inherently less variable than those of Black Noddies. Consequently, if the mark recapture estimator is applied to Black Noddies they

may still be less likely to meet standards than RFBO. Because marking birds may be a significant source of disturbance, we encourage the use of designs that would also allow the survival probabilities of marked birds to be estimated. This will provide an estimate of both the probability of nesting and the probability of survival.

We recommend that detectability of nests be assessed in a similar fashion as Black Noddies (Recommendation 1; Section III). However, we expect that detectability is a minor issue as long as observers can keep track of individual nesting trees. When counting RFBO, we suggest that protocols include methods for keeping track of trees.

Reproductive data. Desired precision (95% confidence interval = 0.1) can be accomplished by sampling fewer plots than are currently sampled. Sampling five plots would satisfy this requirement with 95% confidence (Figs. 2.8 and 2.9). If fewer plots are monitored, we suggest that observers calculate confidence intervals every year and increase plot sizes if 95% confidence intervals are larger than 0.10.

GREAT FRIGATEBIRD (*FREGATA MINOR*)

Data availability

Tern Island. We used 23 years (1982-2004) of MICs for Great Frigatebird (GRFR) breeding pairs across the entire island (Fig. 2.10). No plot-level information regarding reproductive success was available. Average incubation length is 55 days.

Johnston Atoll. We used 13 years of MICs, collected between 1983 and 1995 (Fig. 2.11). No plot-level information regarding reproductive success was available.

Laysan. We used 12 years of data, collected between 1993 and 2004 (Fig. 2.11). Count data for 1996 are missing. These data were not collected with standard MIC methodology; instead the number of breeding pairs across the entire island was counted once each year. Therefore, we expected the Laysan data to have greater observation error than the data from the other two islands. No plot-level information regarding reproductive success was available.

Midway Atoll. We used seven years of data collected between 1985 and 1996 (Fig. 2.11). No plot-level information regarding reproductive success is available.

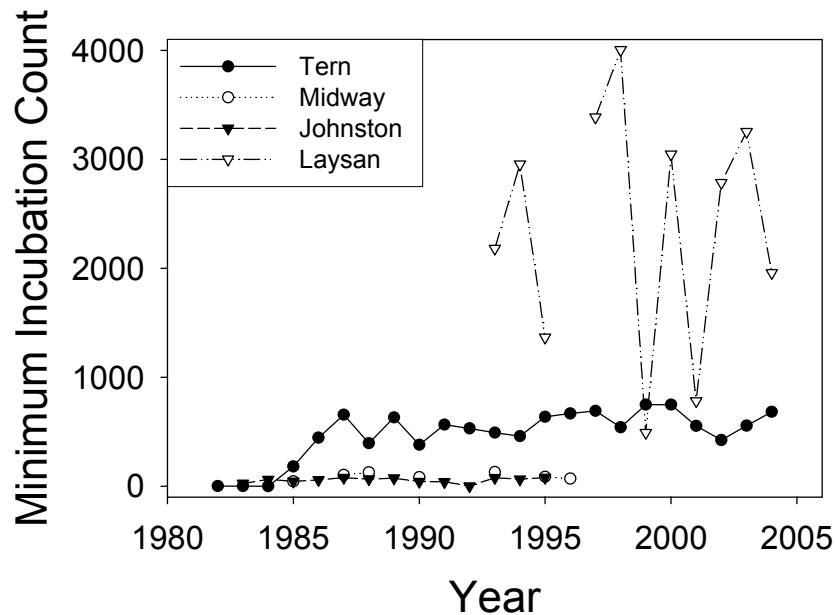


Figure 2.11. GRFR maximum Mean incubation counts (MICs) on Tern, Midway, Johnston, and Laysan. In general, there are approximately between 50 and 150 breeding pairs on Midway and Johnston Atolls. The counts on Tern and Laysan Islands are much larger, but the count on Laysan is highly variable.

Prospective power of MIC data

We found no evidence that variance in population counts increased with population size for any island. Consequently, we did not log transform counts prior to trend analysis. We could detect no serial autocorrelation in the data series for any islands. With current methods, data would have to be collected for 11 years on Tern Island, >50 years on Laysan Island, and 16 years on Johnston Atoll (Table 2.5). The methods used to determine detectable trend within 10 years assume an infinite starting population size and should be considered too optimistic. This is especially true for Johnston Atoll, where the starting count of breeding pairs is only 79. Within a 10 year period, declines greater than 63% for Tern Island and 80% for Laysan Island (Table 2.6) are detectable.

Assessment of current methods

It appears that GRFR populations exhibit different levels of variability on different islands. Low power on Laysan Island is due to a high level of variation coupled with a low number of population counts. Low power on Johnston Atoll is likely due to a low population count, as variance is low (Table 2.5).

Table 2.5. Years of observations required to detect trends in GRFR maximum Mean incubation counts (MICs).

Standard	Years of observations			
	Tern Island (SD=126.67) ^a	Laysan Island N=1959 ^b (SD=1188.46) ^a	Laysan Island N=4000 ^b (SD=1188.46) ^a	Johnston Atoll (SD=23.63) ^a
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.05$; power = 0.9 ^c	11	>50	17	16
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.1$; power = 0.9 ^d	9	>30	15	15

^a Standard deviation on a real scale (i.e., not log transformed).

^b We were initially surprised at how many years were required to detect trends on Laysan. We wanted to determine if starting population size was responsible for low power and conducted two simulations. First, we assumed starting population size was equal to the last available count, 1959; second, we assumed that the starting population size was equal to the maximum observed count, 4000. Although having a larger starting population size does improve power in this scenario, 15-17 years are still required to detect relevant trends.

^c Recommendations of Hatch (2003).

^d Proposed standards for the Alaska Maritime National Wildlife Refuge (G.V. Byrd, pers. comm.).

Table 2.6. Change in GRFR maximum Mean incubation counts (MICs) that is detectable over 10 years. The method used assumes an infinite population size and is therefore not valid for Johnston Atoll (count = 79) and, to a lesser extent, Laysan Island (count = 1959).

Standard	Detectable change over 10 years		
	Tern Island	Laysan Island	Johnston Atoll
$\alpha = 0.05$; power = 0.9	63%	80%	39%
$\alpha = 0.1$; power = 0.9	59%	75%	32%

Alternatives

Power will increase if variation in the count can be decreased. However, observed levels of variation on Tern Island and Johnston Atoll are reasonable and low compared to other species. We do not identify alternative methods for Tern Island, because power is near the required standard. Furthermore, on Tern Island, breeding appears to be only moderately asynchronous (Fig. 2.12). While the mark-recapture estimator will provide estimates that better approximate the true breeding population, we suggest that the application of this experimental estimator is better suited to other species, such as Black Noddy, that are highly asynchronous, or Red-tailed Tropicbird, where banding data are currently collected.

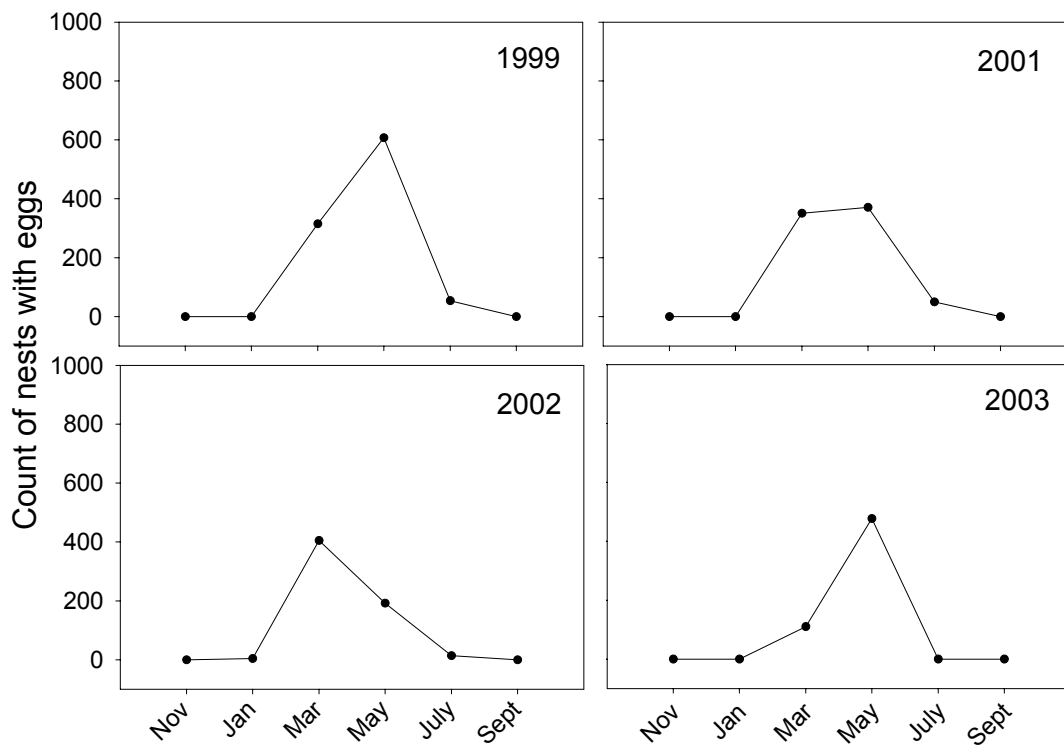


Figure 2.12. GRFR maximum Mean incubation counts (MICs) by month on Tern Island between 1999 and 2003, with the exception of 2000.

We do not identify alternative methods for Johnston Atoll, as low power is likely due to a small population count and use of alternative methods will not improve power. Levels of variation on Johnston were very small (Table 2.6) and cannot be improved upon.

In contrast, variation on Laysan is large. We suspect this is due in part to the fact that GRFR are conducted over just a few days each year. If observers miss peak breeding, counts will vary simply because nests were counted at the wrong time. We suggest that adopting the MIC method on Laysan would reduce count variability.

Precision of reproductive data

No plot level data are currently available for GRFR, so we could not assess the precision of reproductive data. In other words, we could not assess how many plots are necessary for monitoring. If managers choose to monitor breeding success, we suggest using the reproductive success estimator presented in Section IV.

Summary and recommendations

Count data. We conclude that the ability to detect trends in GRFR breeding populations is near standard at Tern Island, and below standard at Johnston Atoll and Laysan Island. Low power at Johnston Atoll is likely due to a low breeding population size and power cannot be improved. Low power on Laysan Island is due to the interaction of high variation and low breeding population size. We suspect that the high variation in count statistics at Laysan Island is mainly due to observers missing peak breeding periods in some years. If counts were conducted throughout the year at regular intervals, we suspect power will improve.

Reproductive data. We suggest applying the estimators for reproductive success generated in Section IV.

LESSER FRIGATEBIRD (*FREGATA ARIEL*)

Data availability

No data are available within the planning area for Lesser Frigatebirds (LEFR). Large colonies occur on Howland and Jarvis Islands (USFWS 2005). However, it is unlikely that the resources will be available to start demographic monitoring. This species would be a good candidate for non-demographic methods (See “Research Priorities” in Section IV); however, these methods would have to be tested extensively.

Summary and recommendations

The average incubation duration of Lesser Frigatebirds is 45 days. Lesser Frigatebirds are believed to renest following failure (USFWS 2005) and likely exhibit aseasonality (and potentially asynchrony). As such, methods for monitoring LEFR will be similar to those recommended for Red-footed Booby and Red-tailed Tropicbird. If there are multiple peaks in breeding within a year, banding will be required to determine if individuals breed multiple times within a year. We suggest basing sampling upon the mark-recapture ratio estimator in Recommendations 8 and 9 in Section III.

We recommend temporarily removing this species as a monitoring priority until monitoring protocols for other more accessible species are solidified and such an endeavor at Howland and Jarvis Islands can be undertaken.

RED-TAILED TROPICBIRD (*PHAETHON RUBRICAUDA*)

Data availability

Tern Island. We used 24 years (1979-2003) of MICs for Red-tailed Tropicbird (RTTR) breeding pairs across the entire island (Fig. 2.13). Average incubation length is 42-46 days. For assessing how many plots are necessary to estimate breeding success we use plot specific data from 1999 to 2002.

Johnston Atoll. We used 12 years (1983-1995) of MICs for RTTR (Figure 2.13).

Laysan Island. While monitoring Common Sandbur (*Cenchrus echinatus*) control plots, observers have conducted MICs of RTTR on Laysan Island. These counts usually cover 1-3 incubation periods. For example, in 1997, three counts were conducted within plots between April and June. The maximum count was only 33 nests and observers concluded only a small portion of RTTR nests occurred in the *Cenchrus* study plots (USFWS 1997). In 1998, two counts were conducted between August and September. The maximum count was 56 nests and observers doubted that all nests in study plots had been found or that the plot counts reflected island totals (USFWS 1998). Therefore, we decided not to use RTTR data from Laysan to quantify variance for power analysis.

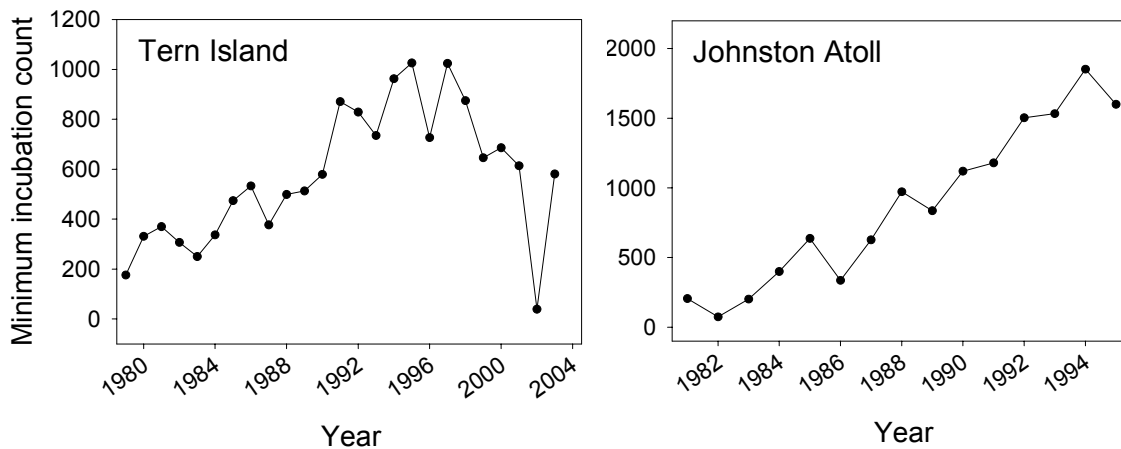


Figure 2.13. Maximum MICs for RTTR on Tern Island (1979-2003) and Johnston Atoll (1982-1995).

Midway Atoll. Population plots have been monitored on Midway Atoll from 2001 to 2004. A sample of birds has been banded within one to three plots. Banding data can be used to avoid double counting re-nesting birds and to avoid problems inherent in MIC methodology. However,

the number of plots monitored has been inconsistent. Three plots were monitored in 2001, two in 2002, and only one in 2003 and 2004. This information may be used in the future to further refine the way MIC methodologies are used for monitoring Red-tailed Tropicbirds.

Reproductive success has been monitored on Midway Atoll from 1998 to 2003. However, the number of plots monitored has varied. In 1998, four plots were monitored, six in 1999, and seven between 2000 and 2001. Then the number of plots decreased to two in 2002 and only one in 2003 and 2004. We used data between 1998 and 2001 (when there were more than two plots monitored) to assess how many plots are necessary to monitor when estimating reproductive success.

Kilauea Point. RTTR are monitored at Kilauea Point. Based upon discussions with Brenda Zaun, it appears that there are only two or three years of sufficient data. We decided this was not a long enough time series to be of use.

Prospective power of MIC data

We found no evidence that variance in population counts increased with population size on Tern Island or Johnston Atoll. Consequently, we did not log transform counts. We could detect no serial autocorrelation in the data series for any islands. With current methods, data would have to be collected for 19 years on Tern Island and seven on Johnston Atoll (Table 2.7). Within a 10 year period, declines greater than 84% on Tern Island and 33% on Johnston Atoll are detectable (Table 2.8).

Table 2.7. Years of observations required to detect trends in RTTR Mean incubation counts.

Standard	Years of observations		
	Tern Island N=581 (SD=184.68) ^a	Tern Island N=1000 (SD=184.68) ^a	Johnston Atoll (SD=151.34) ^a
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.05$; power = 0.9 ^b	19	12	7
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.1$; power = 0.9 ^c	17	11	6

^a Standard deviation on a real scale (i.e., not log transformed).

^b Recommendations of Hatch (2003).

^c Proposed standards for the Alaska Maritime National Wildlife Refuge (G.V. Byrd, pers. comm.).

Table 2.8. Change in RTTR Mean incubation counts that is detectable over 10 years.

Standard	Detectable change over 10 years	
	Tern Island	Johnston Atoll
$\alpha = 0.05$; power = 0.9	84%	33%
$\alpha = 0.1$; power = 0.9	75%	26%

On Tern Island, power was lower than we expected, given the observed variation. We determined that low power was due to a low population size and the type of trend modeled. In 2003, 581 RTTR were counted on Tern Island and this was the starting point for our simulations. When modeling a proportional decrease, the absolute change is small when population size is low. For example, a 50% annual decline results in a large change when the starting population size is 1,000 (a change of 500) versus when the starting population size is only 10 (a change of 5). Because variation does scale with population size in tropicbirds, these small changes are still associated with high variation. When we ran simulations with a starting population size of 1,000, the desired trend was detectable within 12 years (7 years earlier). Because there were large counts of tropicbirds within the dataset (e.g., 1,026 were counted in 1995), we expect that the current monitoring protocol will have sufficient statistical power to detect future population declines.

Assessment of current methods

It appears that RTTR populations exhibit different levels of variability among islands. The Tern Island population is more variable than Johnston Atoll. However, low power on Tern Island appears to be partially a function of small population size. Power will increase if variation in the count can be decreased. We identify a number of alternatives that may increase power in Appendix C. Two alternatives appear promising for RTTRs: (1) Using an alternative metric that is already collected; and (2) because RTTR are asynchronous breeders (Fig. 2.14) combining MIC methods with limited mark-recapture methods on study plots.

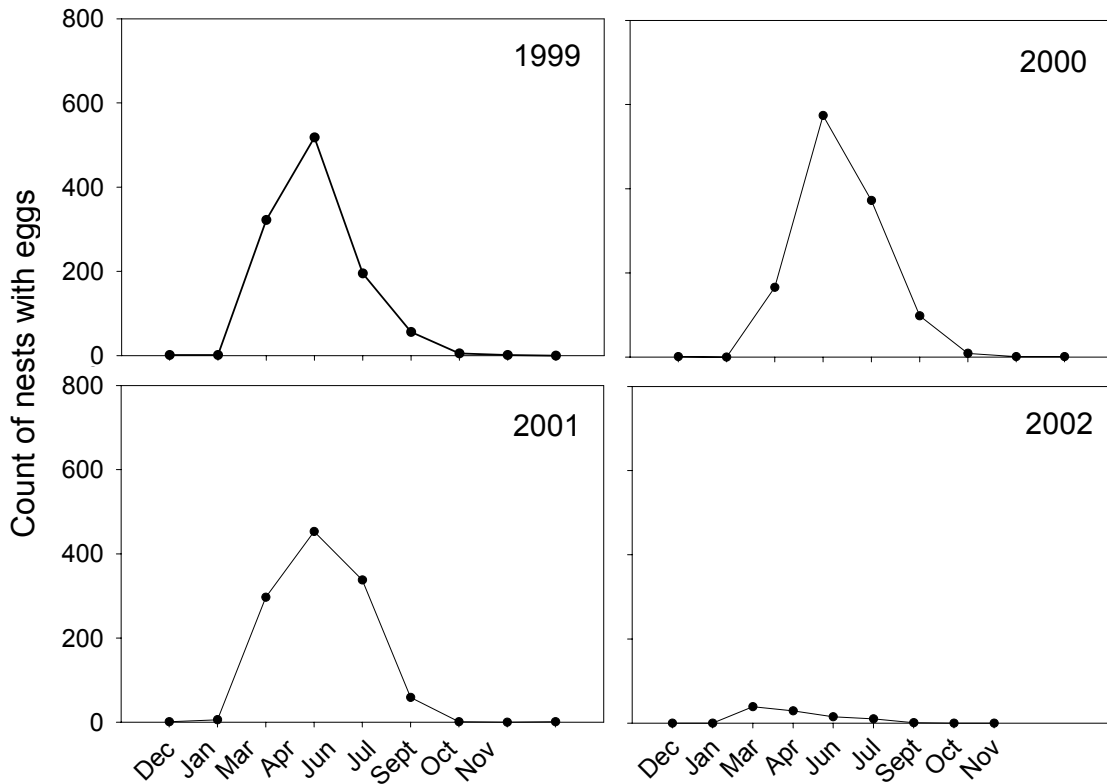


Figure 2.14. Mean incubation counts by month on Tern Island between 1999 and 2002.

Alternatives

Pre-existing metrics. We used data from Tern Island (1999-2003) to compare variance of the MIC with variance of total nests, total nests with eggs, and total nests with chicks. We then calculated the statistical power associated with each metric. All alternative metrics outperform the MIC (Fig. 2.15). In particular, the total number of nests with eggs is much less variable than the MIC, allowing desired trends to be detected within eight years. Apparently, this is because the MIC count is most sensitive to years when a large number of individuals do not breed. For example, breeding failed on Tern Island in 2002. While examining the residuals from the detrended data, we found that MICs varied greatly during 2002, but other metrics did not (Fig. 2.16).

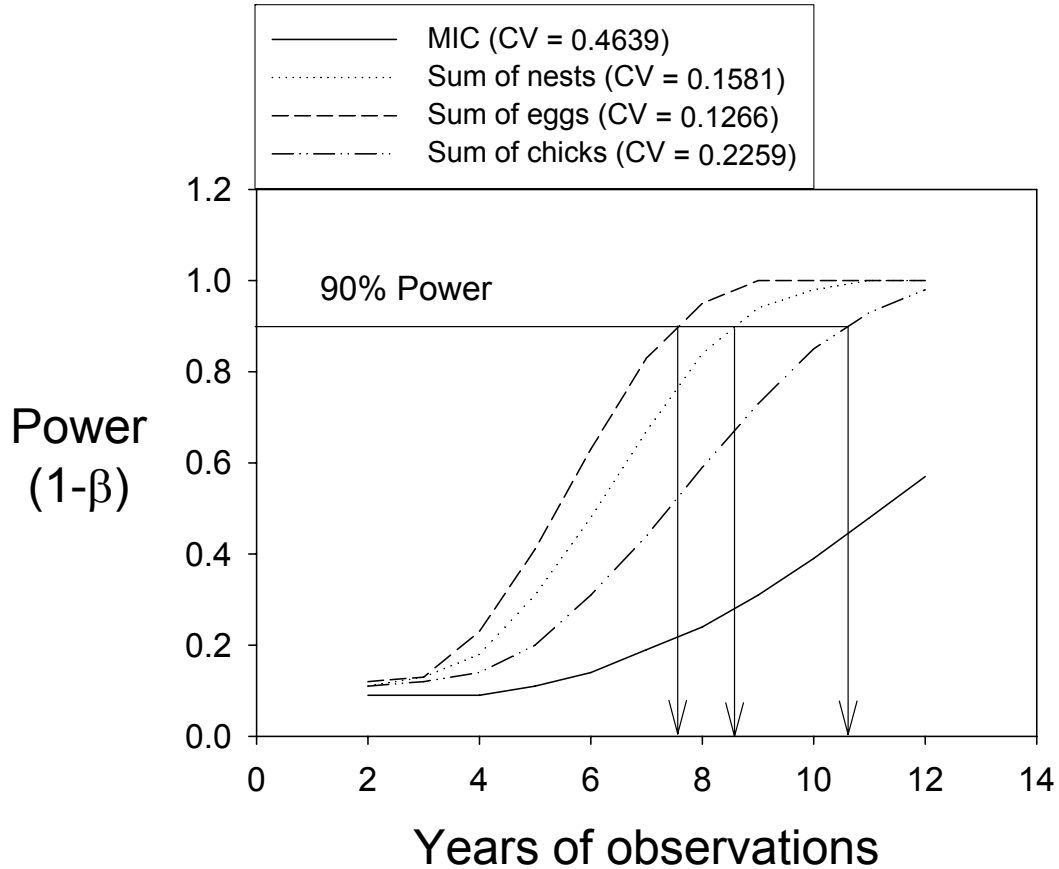


Figure 2.15. Power associated with alternative metrics for population counts for RTTR and a -6.7% annual decline. The coefficient of variation for each metric is presented in the figure legend (CV). Data are from Tern Island between 1999 and 2002. All metrics perform better than MICs. In particular, the sum of nests with eggs shows less variation than MICs and is capable of detecting the desired trend within eight years.

If alternative metrics always exhibit less variation during years when large numbers of individuals do not breed, then we can improve power by using these metrics. However, it is not clear what biological significance underlies this relationship. We have also only observed one year where breeding has almost completely failed (2002). Hence, this conclusion is based upon a single event in the RTTR data. We conclude that the use of alternative metrics be used with caution, but that they may prove to be more effective than MICs. The performance of these metrics must be assessed when breeding fails.

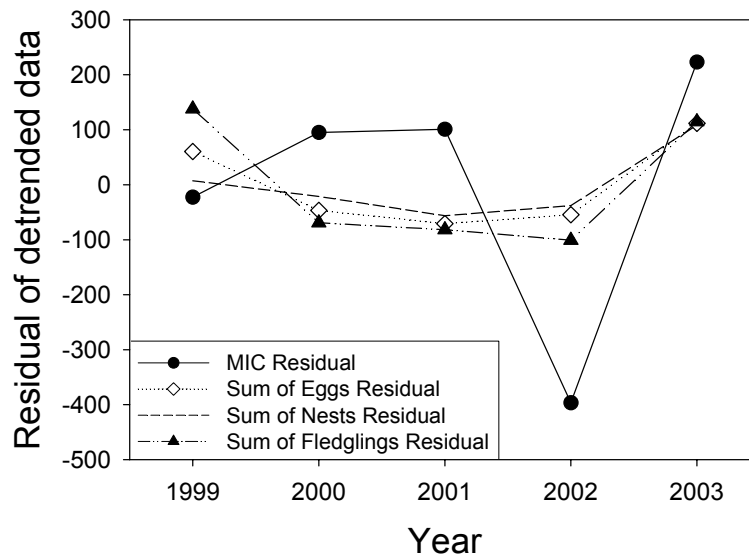


Figure 2.16. Residuals from detrended data used to assess power in RTTR Figure 3. This figure indicates that during 2002, when breeding success was very low, the sum of nest with eggs, sum of all nests, and the sum of fledged nests all exhibited lower variation than the MIC. Years when a large proportion of the population does not breed introduce much variation into MIC data and lead to low prospective power. If alternative metrics are less variable when individuals do not breed, then they may be worth investigating further.

Mark-recapture on study plots. This method is described in detail for Black Noddies and in Appendix C. Because RTTR are moderately asynchronous breeders (Fig. 2.14), combining mark-recapture with MIC methodology is expected to decrease variance. However, there is insufficient pilot data for testing this method. Hence, we do not know how large the increase in power will be.

Precision of reproductive data

Our simulations indicate that fewer plots can be sampled with little increase in bias. If only five plots were sampled, bias would increase by approximately 2% on Tern Island (Fig. 2.17) and 3% on Midway Atoll (Fig. 2.18). The 95% confidence limit for this increase is approximately 6% on both islands. This level of bias is trivial.

However, sampling fewer plots does sacrifice precision. Although the average level of precision does not increase so that confidence intervals are greater than ± 0.1 , the 95% confidence interval for precision rapidly increases beyond 0.1. This suggests that decreasing sampling effort may result in precision that is highly variable. This is true for both Tern Island

and Midway Atoll. To keep confidence intervals less than ± 0.1 with 95% confidence, a minimum of eight plots must be sampled on Tern Island (Fig. 2.17) and seven plots must be sampled on Midway Atoll (Fig. 2.18).

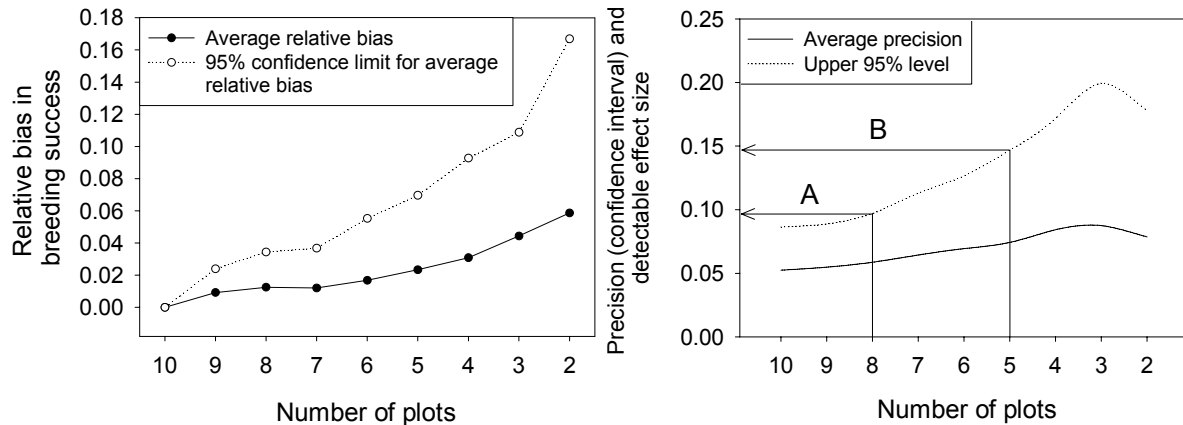


Figure 2.17. Average and 95% confidence intervals of relative bias and precision versus the number of plots on which RTTR reproductive data are collected on Tern Island. All values are in relation to a total of 10 plots. For bias, reducing number of plots from 10 to 5 would result in breeding success being biased by 0.015 on average with a 95% confidence limit of 0.060. The somewhat erratic shape of the curves is caused by having few years of data (1999-2004) and not a small number of simulations (there are 48 simulations for each year and a total of 288 simulations across the entire time period). For precision, two drop lines are shown. In case A, decreasing the number of plots from 10 to 8 results in a minimum detectable effect size of 0.08 with 95% confidence. In case B, decreasing the number of plots to five results in a minimum detectable effect size of approximately 0.15.

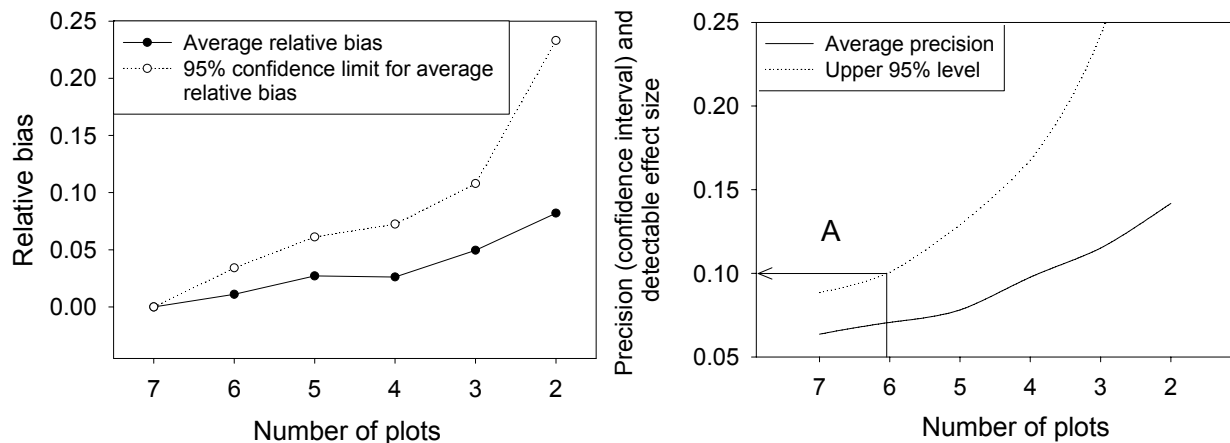


Figure 2.18. Average and 95% confidence interval of relative bias and precision versus the number of plots on which RTTR reproductive data are collected on Midway Atoll. All values are in relation to a total of seven plots. For bias, reducing number of plots from seven to five would result in breeding success being biased by 0.03 on average with a 95% confidence limit of 0.06. The somewhat erratic shape of the curves is caused by having few years of data (1997-2001) and not a small number of simulations (there are 48 simulations for each year and a total of 288 simulations across the entire time period). For precision, one drop line (A) is shown. To maintain a confidence interval of 0.1 with 95% confidence, more than six plots are necessary.

Summary and recommendations

Count data. We conclude that the ability to detect trends in RTTR breeding populations is above standard at Johnston Atoll and below standard at Tern Island. We suggest that investigating alternative metrics is warranted. The use of total nest with eggs as a metric drastically improved power, but this is primarily because the MIC for 2002 was exceedingly low. Because few instances with such low breeding success have been documented, we are hesitant to recommend this metric as a solution. More data are needed to quantify the relationship between low breeding success and variation in total nests with eggs.

We also suggest that MICs be combined with mark-recapture (i.e., the mark-recapture ratio estimator) on study plots (See description in Appendix C and account for the Black Noddy). Red-tailed Tropicbirds may provide a unique opportunity to test the mark-recapture ratio estimator, because mark-recapture studies are ongoing on Tern Island and Midway Atoll. With existing data from Midway Atoll, the ability of a mark-recapture approach to correcting MICs could be evaluated. On Tern Island, Dr. Ian Jones currently bands RTTR, but does not keep track of nesting behavior of marked individuals. We recommend that USFWS consider collaboration with Dr. Jones to test the mark-recapture estimator on Tern Island. On Midway Atoll, mark-recapture techniques are used on one to three study plots. We suggest that the number of plots on Midway be increased to five if resources allow. We also suggest that MIC methods be used on Midway and that the mark-recapture information be used to adjust MICs to account for asynchrony.

Reproductive data. For Tern Island, the desired precision (95% confidence interval = 0.1) for breeding success can be accomplished by sampling fewer plots than are currently sampled. Sampling eight plots would satisfy this requirement with 95% confidence (Figs. 2.17 and 2.18). Similarly, on Midway Atoll, monitoring more than six plots would achieve desired precision. If fewer plots are monitored, we suggest that observers calculate confidence intervals every year and increase the number of plots if 95% confidence intervals are larger than 0.10. For a more general description of how to estimate breeding success and a discussion plot number/size issues, see Section IV.

SOOTY TERN (*STERNA FUSCATA*)

Data availability

No data are available within the planning area. However, Sooty Terns (SOTE) are present on most of the Northwestern Hawaiian Islands, and colonies also exist on Howland Island, Baker Island, Jarvis Island, Palmyra Atoll, and Johnston Atoll. Average incubation duration is 30 days. In some locations, SOTE are sensitive to disturbance, especially during the early stages of nesting when thermal stress or egg predation may cause nest loss (E. Flint, pers. comm.). Sooty Terns also nest at densities with multiple nests per square meter. As such, moving about the colony to collect either count or reproductive data will prove difficult.

Potential methods

We suggest that SOTE will require the development and application of new monitoring methods. A possible solution is to monitor nest density and reproductive success with pole-mounted video camera or digital camera with an interval timer (e.g., Harding et al. 2005). Palmyra would be an excellent location to try this method, as there are pre-existing trees or poles adjacent to, or within, all colonies. The proposed method is as follows:

- 1) Before the breeding season begins, the approximate boundaries of the colony should be mapped. This mapping could be accomplished by marking the colony boundaries with rocks during the nesting season and then recording these locations with a GPS after breeding is completed. Depending on local conditions, some effort may need to be invested in the best way to delineate the colony boundaries.
- 2) Plots within the colony that can be monitored with a pole-mounted camera should be selected and marked (Fig. 2.19). Plots will have to be relatively small, but attaining a sample of >20 birds per plot should be sufficient. Plot boundaries should be marked in such a manner that they will be visible in the recorded images.
- 3) Poles are erected adjacent to the plots. We envision digging a series of holes, with one hole near each video plot. Holes could have a larger diameter pipe in them, providing a sleeve for the pole, and could be covered when not in use. If holes cannot be dug, then a moveable base plate may hold the pole erect (Fig. 2.20). One camera on one pole can then be moved around to sample the different plots. A large diameter PVC pole may help minimize strikes, but the

number of strikes will need to be monitored. During breeding periods the density of nesting birds and the number of nestlings and potentially eggs can be quantified on plots by attaching a camera to a pole (Fig. 2.19).

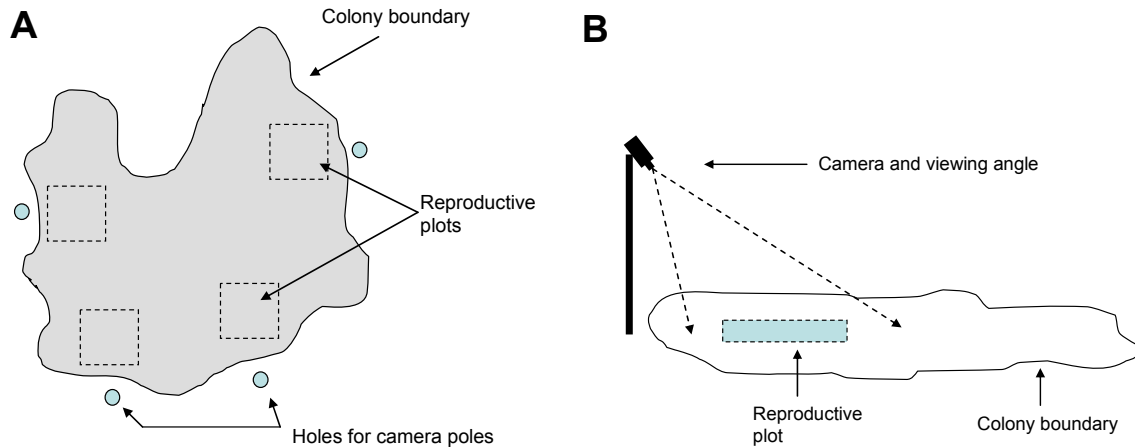


Figure 2.19. A proposed set-up for monitoring Sooty Tern plots with video cameras. (A) View of a hypothetical colony (gray shape), monitoring plots marked with dotted lines, and holes for camera poles marked as circles. (B) Horizontal view of a monitoring plot (dotted square) showing the viewing angle of the camera on a PVC pole. One camera pole can be moved from hole to hole along the colony edge.

4) Once filming is complete, the pole can be removed and the images can be played on a computer screen. Nest locations could be mapped either by attaching a piece of transparent film to the computer screen and then marking the nest location and status of nests on the film, or simply by marking the digital image with available image software.

Details about how to attach the camera to the pole, where to position the camera relative to the plot, how long plots should be filmed, and how to classify nest status will need to be resolved. Furthermore, the potential for birds to injure themselves by striking the pole should be considered. Estimators for density and reproductive success will be those for a simple random sample (see Recommendation 4, Section III).

Another potential method for this species would investigate the variability in nesting densities. Some species of tern, such as the Elegant Tern (*Sterna elegans*), have low variability in nesting densities. If this is true for SOTE, then breeding population size could be estimated by measuring the area of the colony in each year. This could be accomplished by walking around the edge of the colony with a GPS unit on continuous readout. SOTE may be asynchronous

across an island, but are typically synchronous within sub-colonies. If studies can verify nesting density is homogenous, then this method could be a reliable and quick method for estimating the number of SOTE breeding pairs.

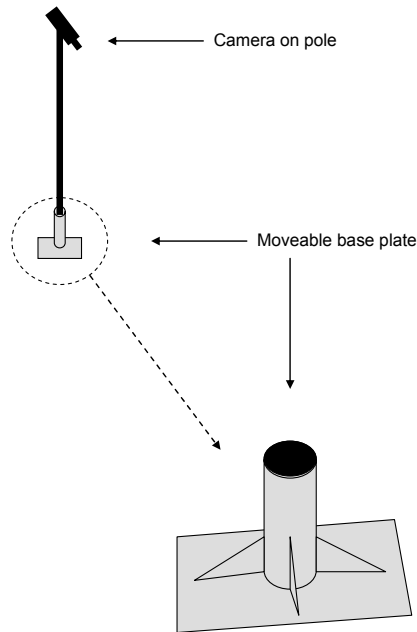


Figure 2.20. A proposed method for placing poles near Sooty Tern plots: a platform with a hollow tube, within which a camera pole can be placed. The platform could be stabilized with sandbags or rocks.

SOTE would also be a candidate for physiological methods (Section V), but this will require studies of known age (i.e., banded) birds to verify and calibrate patterns. The potential sensitivity of SOTE to disturbance may make banding studies problematic on some islands.

WEDGE-TAILED SHEARWATER (*PUFFINUS PACIFICUS*)

Data availability

Data for Wedge-tailed Shearwaters (WTSH) are collected at a number of refuges. At no location has WTSH data collection been a priority. Consequently, data collection has been erratic and potentially unreliable. In response to inquiries regarding data quality, Beth Flint (USFWS, Honolulu) suggested that WTSH sampling and monitoring be treated as a “blank slate.”

Sampling design

The average incubation duration of Wedge-tailed Shearwaters is 54 days. Instead of suggesting a census of all burrows, we present a sampling design that will be generally applicable to any accessible WTSH colony (and other species of burrow nesters). Even if the entire colony could be enumerated, a sampling design will reduce disturbance across the colony and save time and effort. We recommend monitoring plots rather than using distance or transect-based sampling. To determine activity, burrows will likely have to be visited a number of times. Using plots will allow researchers to mark individual burrows and limit disturbance across the colony. This design is detailed below:

- 1. Map the colony boundaries.* The boundaries of the colony need to be mapped and colony area should be calculated. This can be accomplished within GIS. We recommend that general colony boundaries be verified on a yearly basis.
- 2. Randomly locate plots within the colony.* Within the colony, plots should be randomly located. There are two methods for random plot assignment, depending upon whether or not there is an established grid of possible sampling plots (hereafter sampling frame) across the colony (Fig. 2.21). If the sampling frame is defined, then a random sample of plots is selected. If the sampling frame is not well defined (i.e., if there is not an established grid over the colony), then random x and y coordinates are chosen, rather than plots. In theory, statistical estimators rely on the random selection of plots, not x-y coordinates. However, the biases are negligible in practice. Especially in real situations (Fig. 2.21), where colony boundaries may not conform to plot shapes, the random selection of x-y coordinates is preferred. The critical assumption is that the plots are representative of the colony as a whole. If this is true, then our estimators will be valid. Once plots are randomly selected, we recommend resampling the same plots each year.

Some authors suggest sampling different plots each year. While sampling a new sample of plots each year will provide more statistical power to estimate true abundance of burrows, it provides less power to detect trends in abundance or reproductive success. This occurs because by sampling new plots we are adding an additional source of “noise” (variation between new plots) into analyses of trends (see Thompson et al. 1998; page 177, “Plot reselection”). Make certain that the plots remain representative of the colony as a whole.

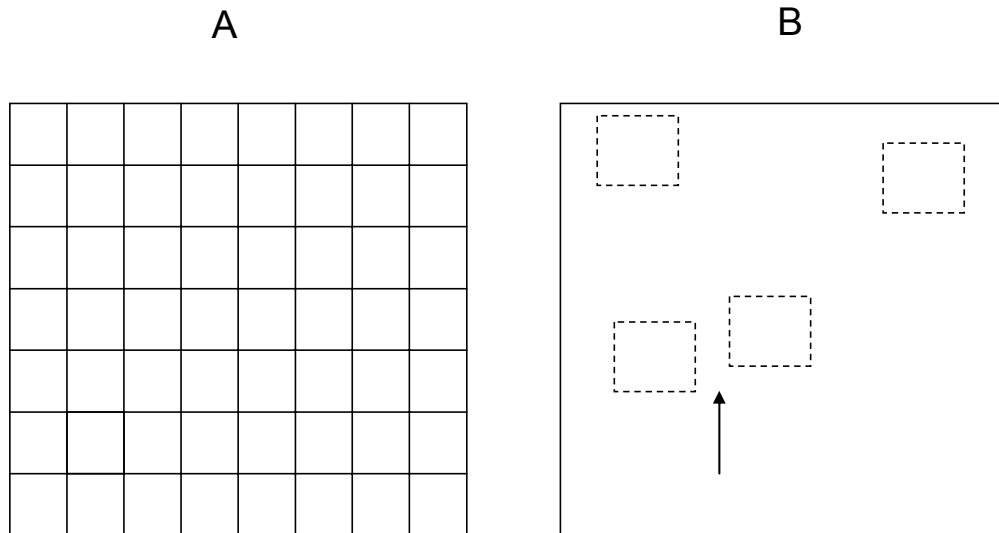


Figure 2.21. A well defined (A) and an “open” (B) sampling frame. In instance A, the colony is gridded, each possible plot is numbered and then plots are randomly selected. In instance B, random points are selected as plot centers. With the open sampling frame, some plots are not possible (e.g., there cannot be a plot where the arrow indicates in instance B).

3. Exceptions to random plot location. If nests are clustered, we suggest recording nest locations with a GPS and then defining clusters. Instead of selecting random plot locations, center plots on clusters of nests and then randomly select nest clusters (Fig. 2.22). If plots are already pre-existing and refuge biologists want to monitor these plots for logistical reasons, we suggest that biologists very carefully consider if the existing plots are representative of the colony as a whole. If pre-existing plots are not representative of the colony, randomly select new plots. The required plot sizes and the number of plots will depend upon the variation between plots in burrow abundance and nest success. When plots are more variable, researchers will need to sample a greater number to achieve the same precision as when they are less variable. As a start, we recommend having a minimum of eight plots that are big enough so that there are more than

20 active burrows in each plot. When burrows are clustered, a stratified random sampling design may be appropriate. In this approach, high density areas would be identified, and then a random sample of these areas selected (Fig. 2.22). However, such a strategy should be employed cautiously if there is any reason to expect that differences between high and low density areas are important.

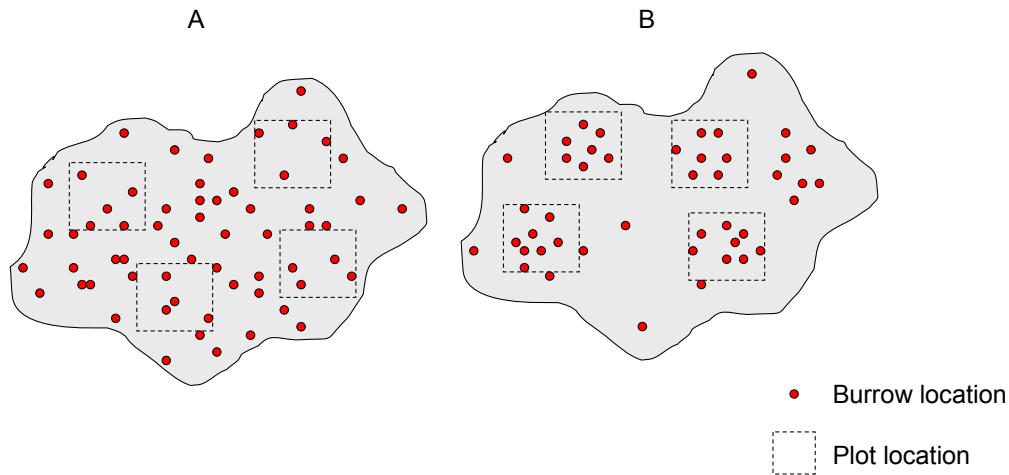


Figure 2.22. A sampling frame where nests are randomly distributed (A) and where nest locations are clustered (B). When burrows are randomly distributed, plots can be centered on random points. When burrows are clustered, plots can be placed on clusters and then random clusters are selected as plot locations. More than one plot may occur on a cluster but plots should be non-overlapping. For abundance, we suggest sampling enough plots that the coefficient of variation of the number of nests per plot is less than approximately 10% (the spreadsheet will calculate this for you). For reproductive success, we recommend that you sample enough plots to maintain a 95% confidence bound of less than 0.1. The attached Excel spreadsheets will calculate confidence intervals and bounds.

4. *Monitoring abundance.* Mark and label all burrows (to keep track of them). On plots where abundance is quantified, we suggest monitoring activity in a subset of nests to determine when incubation has begun. Unless burrows are shallow, we suggest using a burrowscope (e.g., Dyer and Hill 1991) to verify activity. If burrowscopes are used, more than one should be on hand such that data are not lost if a single piece of equipment fails. Wedge-tailed Shearwaters are very synchronous and all pairs should start incubating within approximately two weeks of when the first egg is observed. We suggest monitoring burrows once each week on abundance plots. Once an egg is detected in a burrow, we suggest that the burrow no longer be monitored. In other words, the goal is to determine the number of active burrows on a plot. Once an egg is detected, we are confident the burrow is active in that year and there is no need to disturb the

burrow further. We suggest ceasing to monitor abundance plots two weeks after the last new egg is detected (unless the plot is also used to estimate breeding success).

5. *Monitoring reproductive success.* We suspect that variation in burrow density between plots will be greater than variation in breeding success. As such, it may be necessary to monitor more plots to assess abundance than to assess breeding success. If this is true, then we suggest monitoring a subset of the abundance plots when assessing breeding success. On plots where breeding success is quantified, we suggest monitoring nest stage once each week until chicks fledge from the nest. Determining fledging will be somewhat subjective. The published range of variation in age at fledging ranges from 99 days on Kure Atoll (Woodward 1972) to 115 days at Kilauea Point (Byrd et al. 1983). Visual cues may also be useful. Fledging chicks will look very similar to adults, but with traces of down (Whittow 1997). We suggest that chicks less than approximately 100 days old (i.e., 100 days post-hatch) probably should not be considered to have fledged successfully. We leave determination of fledging to the local biologist, but emphasize that whatever criterion is used needs to be documented. Unless there is a desire to estimate survival rates, we recommend that birds not be banded. Estimation of survival will require a long-term and focused effort. Banding WTSH is likely not necessary to determine trends in the abundance of breeding pairs or reproductive success and likely causes disturbance.

6. *Data analysis.* See Section III for information on selecting an appropriate estimator. We suspect that Recommendation 4 (in Section III) will be useful if burrows can be counted without error on plots (See Recommendation 1; Section III). For abundance, we suggest sampling enough plots that the coefficient of variation is less than approximately 10% (the spreadsheet will calculate this for you). For reproductive success, we recommend that you sample enough plots to maintain a 95% confidence bound of less than 0.1. However, you may use what limits you prefer (the attached Excel spreadsheets will work with any limits). Note that hatching success ($\#hatched/\#eggs$) can be estimated with the estimator for reproductive success by substituting $\#hatched$ for $\#fledged$. The methods of estimation are the same. Formal analyses of trends can take many forms. In the simplest case, trends can be visually assessed by plotting abundance and the associated 95% confidence intervals over time. This will be sufficient for management uses, as long as observation error is low (i.e., as long as error bars are small). As data accumulate, we suggest a formal analysis of trend using generalized linear models and

accounting for autocorrelation if necessary (e.g., using generalized estimating equations). Other, more complicated, options are available. These can be assessed at a future date, if plot level data are recorded and archived each year. We strongly suggest that all plot level data are recorded and archived. We recommend saving a copy of the attached spreadsheets for each breeding season. If the plot level data are archived, these data can be applied to future trend analyses.

CHRISTMAS SHEARWATER (*PUFFINUS NATIVITATIS*)

Data availability

Tern Island. There are two series of count data available for Tern Island. Counts are made by canvassing the entire island. The Pacific Seabird Monitoring Database contains consistent records from 1989–1996 and data files from Tern Island have records covering 1991 to 2002. Unfortunately, the files do not indicate the same number of individuals within overlapping years (Fig. 2.23). However, both files indicate that population sizes are low (Range: 11–40 breeding pairs) and the resulting variances are similar (SD presented in Fig. 2.23). Breeding success is also measured on Tern Island. The number of fledged chicks is divided by the number of active nests to yield breeding success. Reproductive success is not sampled on plots, but is calculated as an island-wide census.

Data from other islands. Limited mark-recapture studies of CHSH are occurring on Kure Atoll (e.g., Vanderlip and Marie 2003; Vanderlip 2004). Individuals are banded upon the runway at night, but nests are not monitored. Although these data are of value, opportunistic banding cannot address counts of breeding birds nor reproductive success without additional monitoring. Reproductive plots were monitored on Midway Atoll between 1995 and 2001. John Klavitter (Biologist, Midway Atoll National Wildlife Refuge) provided us with raw data files from Midway Atoll. However, because we did not have access to detailed meta-data, we had difficulty determining how to interpret the CHSH data in the Midway data files and did not use them for this analysis.

Prospective power of MIC data

Average incubation length is 52 days. There was no evidence that variance increased with population size and it appears that population size is uniformly low. Therefore, log transformations of counts are inappropriate and we assume that variance is constant. We could not detect temporal autocorrelation in either data series. Due to low population size, assumptions regarding variance structure are critical. Depending upon the starting population size, the type of trend (linear or multiplicative), and how variance relates to the population mean (constant variance or constant coefficient of variation), the number of years required to detect a 50% decline with $\alpha = 0.05$ and 90% power will vary from five years to infinity (Table 2.8). In other words, the assumptions we make become critical for assessing prospective power when

populations are less than approximately 50 individuals. The most recent count available to us suggest that there were 11 breeding CHSH. Therefore, we cannot reliably estimate prospective power for CHSH.

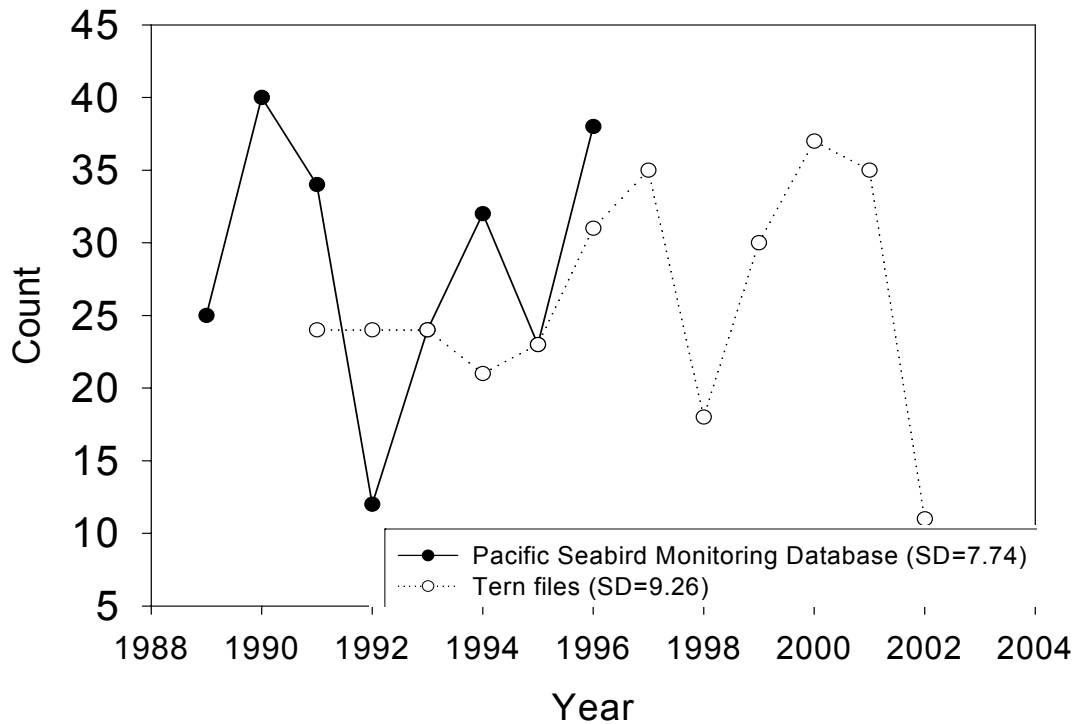


Figure 2.23. Count data for CHSH from Tern Island.

Assessment of current methods

It appears that CHSH populations are stable on Tern Island. While the relevant trend cannot be detected within the specified time-frame, this is mainly due to low population size and not due to high levels of variation. Essentially, the limiting factor for detecting trends is population size, not how the population is monitored (as long as counts are generally accurate).

Alternatives

Because CHSH are not known to renest following nest failure and lay only one egg (Seto 2001), we expect that counts will be strongly correlated with the abundance of breeding birds (as long as detectability is high). Using alternative metrics will not improve prospective power and the population is too small for mark-recapture methods to provide useful information. If the entire island can be counted and detectability is high (See Recommendation 1; Section III), a

simple count may result in a census. If the entire island cannot be sampled, but detectability is high, we suggest using the simple random sample estimator (Recommendation 4; Section III).

Table 2.8. Years of observations required to detect trends in CHSH counts on Tern Island. We only present the recommendations of Hatch (2003) to illustrate the range of variation in prospective power with low population sizes. See Appendix A for a detailed explanation of simulation methods.

Starting population size	Years of observations			
	Constant CV		Constant variance	
	Multiplicative trend	Additive trend	Multiplicative trend	Additive trend
10	14	5	Infinity	16
15	14	6	Infinity	15
20	14	8	30	14
25	14	9	23	14
30	14	10	19	14
35	14	12	16	14
40	14	13	14	14
45	14	14	14	14
50	14	14	14	14

Precision of reproductive data

No plot level data are currently available for CHSH, so we could not assess the precision of reproductive data. Variation in reproductive success can be quantified via plot studies, but population sizes are likely too low for standard sampling designs to be of much use. We suggest the continuation of current methods for CHSH. If high density colonies are monitored in the future, we suggest using the reproductive success estimator presented in Section IV.

Summary and recommendations

Count data. For CHSH on Tern Island, the limiting factor for detecting trends is population size, not how the population is monitored (as long as counts are generally accurate). We do not think that improving sampling designs for CHSH will improve power. If managers are interested in population trends, we recommend current methods are continued with the understanding that data analysis will be qualitative. In other words, the interpretation of count data will rely on non-statistical measures. Whether or not such data justify the disturbance created by canvassing the island for CHSH nests should be considered by refuge biologists. Furthermore, we do not think that banding individuals on Tern should be considered a high priority. Sample sizes are

definitely too low for estimating survival with banded individuals.

On other islands, where CHSH are likely more abundant, we suggest using the estimators for a simple random sample (see Section III). Possible locations include: Midway Atoll, Laysan Island, and Kure Atoll. Data files indicate that there are significant populations of CHSH on all these islands. In particular, Fefer et al. (1984) estimated that there were between 1,500 and 2,000 breeding pairs on Laysan Island and small populations (20-50 pairs) on Midway and Kure. However, with the recent eradication of rats from Midway and Kure, the breeding populations of Christmas Shearwaters have been growing, and these locations are now logistically feasible for monitoring.

Reproductive data. Estimates of breeding success from Tern Island are based upon a census. Hence, there is no variation in breeding success. Variation could be quantified via plot studies, but population sizes are likely too low for standard sampling designs to be of much use. We suggest the continuation of current methods for CHSH. If managers are interested in reproductive trends, we recommend current methods are continued with the understanding that data analysis will be qualitative. In other words, the interpretation of count data will rely on non-statistical measures. If high density colonies are monitored (e.g., Midway Atoll, Laysan Island, and Kure Atoll) we suggest using the estimator in Section IV.

A special note on Kure Atoll. Individuals are currently banded on Kure Atoll. Without a dedicated investigation, we cannot definitely state that the current level of banding effort is worthwhile. The required sample size will depend upon many factors, such as the probability of encounter (i.e., what is the likelihood that a band is observed given the birds is present and breeding), variability in apparent survival, etc. No quantitative estimates of CHSH survival are available (Seto 2001). Because so little is known about this species, we suggest that monitoring and/or banding studies continue and be expanded at Kure Atoll. We also suggest that population counts and reproductive success be monitored on plots at Kure. In particular, we recommend that reproductive success be monitored at Kure, as reproductive success is expected to reflect food availability. Kure is at the end of the Northwestern Hawaiian Island chain and, when combined with monitoring at Tern Island, monitoring at Kure will provide information regarding food resources over a large geographic area.

BULWER'S PETREL (*BULWERIA BULWERII*)

Data availability

Tern Island. There is a long data series of count data available from Tern Island. This data series spans from 1981 to 2004, but there were significant data gaps between 1981 and 1986 and in 1988. We use only the continuous series from 1989 to 2004 (Fig. 2.24). Breeding success (total chicks fledged per total eggs) is also measured on Tern Island and Johnston Atoll. The number of fledged chicks is divided by the number of active nests to yield breeding success. Plot level information is not available for either Johnston Atoll or Tern Island.

Data are opportunistically collected for Bulwer's Petrel (BUPE) on Kure Atoll (Vanderlip 2004). It appears that BUPE may be trying to recolonize Green Island at Kure, but sample sizes are very small. Three breeding pairs were observed in 2002, 0 in 2003, and 3 in 2004. There are too few data for quantitative analyses.

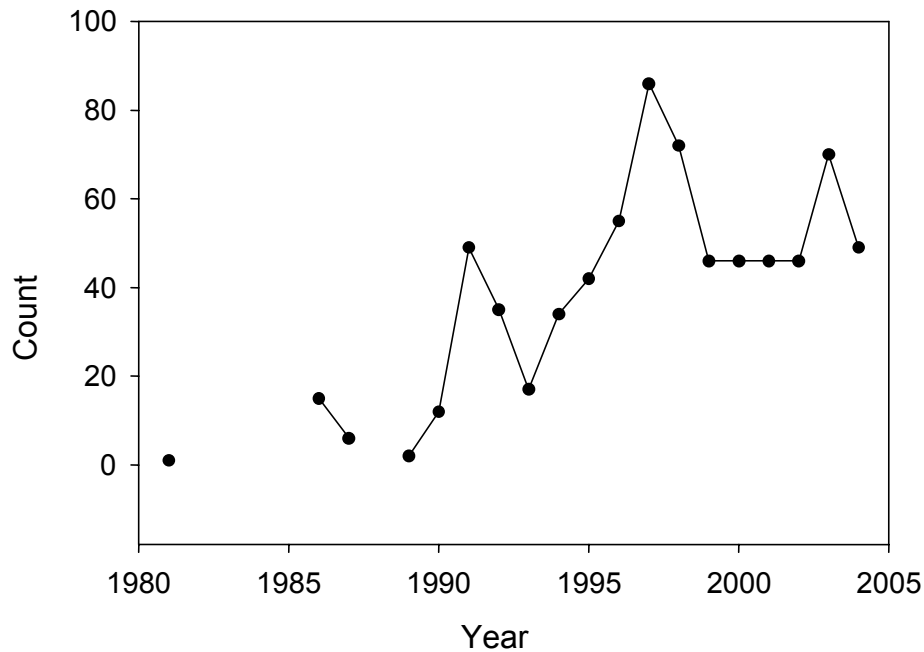


Figure 2.24. Count data for BUPE on Tern Island. Only the continuous series between 1989 and 2004 was used to determine the variance for power analyses.

Prospective power of MIC data

The average length of incubation is 44 days. For Tern Island, there was no evidence that variance increased with population size and it appears that population size is uniformly low. Therefore, log transformations of counts are inappropriate and we assume that variance is constant. We did not detect temporal autocorrelation in the data series and assume observations are independent. Due to low population size, assumptions regarding variance structure are critical. Depending upon the starting population size, the type of trend (linear or multiplicative), and how variance relates to the population mean (constant variance or constant coefficient of variation), the number of years required to detect a 50% decline with $\alpha = 0.05$ and 90% power will vary from six years to infinity (Table 2.9). In other words, the assumptions we make become critical for assessing prospective power when populations are less than approximately 50 individuals. The most recent count available to us suggest that there were 49 breeding BUPE. Therefore, we cannot reliably estimate prospective power for BUPE on Tern Island.

Assessment of current methods

Although the relevant trend cannot be detected within the specified time-frame, this is mainly due to low population size and not due to high levels of variation. Essentially, the limiting factor for detecting trends is population size, not how the population is monitored (as long as counts are generally accurate). Similar to Christmas Shearwater, we cannot reliably estimate prospective power for BUPE.

Alternatives

Because BUPE are not known to re-nest following nest failure and lay only one egg (See review in Megyesi and O'Daniel 1997), we expect that counts will be strongly correlated with the abundance of breeding birds (as long as detectability is high). Using alternative metrics will not improve prospective power and the population is too small for mark-recapture methods to provide useful information. If the entire island can be counted and detectability is high (See Recommendation 1; Section III), a simple count may result in a census. If the entire island cannot be sampled, but detectability is high, we suggest using the simple random sample estimator (see Recommendation 4, Section III).

Table 2.9. Years of observations required to detect trends in BUPE counts on Tern Island. We only present the recommendations of Hatch (2003) to illustrate the range of variation in prospective power with low population sizes.

Starting population size	Years of observations			
	Constant CV		Constant variance	
	Multiplicative trend	Additive trend	Multiplicative trend	Additive trend
10	15	6	Infinity	23
15	15	6	Infinity	20
20	15	7	Infinity	19
25	15	10	Infinity	19
30	15	10	Infinity	18
35	15	11	35	18
40	15	13	27	18
45	15	13	23	18
50	15	14	20	18

Precision of reproductive data

No plot level data are currently available for BUPE, so we could not assess the precision of reproductive data. In other words, we could not assess how many plots are necessary for monitoring. However, although variation in reproductive success can be quantified via plot studies, population sizes are likely too low (< 50 breeding pairs) for standard sampling designs to be of much use. Unless observers are missing significant numbers of nests, we suggest continuing current methods for BUPE. If managers are interested in reproductive trends, we recommend current methods are continued with the understanding that data analysis will be qualitative. In other words, the interpretation of count data will rely on non-statistical measures. If high density colonies are monitored in the future, we suggest using the reproductive success estimator presented in Section IV.

Summary and recommendations

Count data. For BUPE on Tern Island, the limiting factor for detecting trends is population size, not how the population is monitored (as long as counts are generally accurate). We do not think that improving sampling designs for BUPE will improve power. If managers are interested in BUPE trends, we recommend current methods are continued with the understanding that data analysis may have to be qualitative. If BUPE are monitored where they are more abundant, such

as Laysan, we suggest using the estimators for a simple random sample (see Section III).

Reproductive data. Estimates of breeding success from Tern Island are based upon a census. Hence, there is no variation in breeding success. Variation could be quantified via plot studies, but population sizes are likely too low for standard sampling designs to be of much use. We suggest continuing current methods for BUPE with the understanding that data analysis will be qualitative in these locations. If high density colonies are monitored we suggest using the estimator in Section IV.

BONIN PETREL (*PTERODROMA HYPOLEUCA*)

Data availability

Few data are available within the planning area for Bonin Petrels (BOPE). Scant records exist for both Tern Island and Midway Atoll, thus there are not enough data to parameterize analyses for assessing methods. However, BOPE breed from Tern Island to Kure Atoll and large breeding colonies exist within the Northwestern Hawaiian Islands. Over 100,000 pairs are believed to breed on Lisianski (USFWS 2005), 50,000-75,000 pairs on Laysan (USFWS 2005), and >70,000 pairs on Midway Atoll (Small 1999). Average length of incubation is 49 days.

Recommendations

Although BOPE are not currently monitored intensively, we recommend that the final monitoring plan accommodate this species. First, like Tristram's Storm-Petrel (*Oceanodroma tristrami*) and Bulwer's Petrel over 50% of the known population of BOPE breed within the planning area. Second, unlike Tristram's Storm-Petrel and Bulwer's Petrel, quantification of BOPE trends will not be limited by small population size. Third, BOPEs are synchronous breeders, lay only one egg per year, and are not known to reneest (Seto and O'Daniel 1999; USFWS 2005), so MIC methods are useful, and mark-recapture estimators are not necessary for estimating the number of breeding pairs. However, note that mark-recapture estimators will yield information that simple counts will not, such as the likelihood of skipping breeding seasons, probability of encounter, total population size (breeders and non-breeders), and survival. Hence, BOPE are a species of management concern that can logistically be monitored with statistical rigor. If burrows are dense enough that a minimum of 20 nests can fall within plots, we suggest using the simple random sample estimator (see Recommendation 4, Section III) and the estimator for reproductive success presented in Section IV. For more information on possible sampling designs, also see the section on Wedge-Tailed Shearwaters.

TRISTRAM'S STORM-PETREL (*OCEANODROMA TRISTRAMI*)

Data availability

There are no data available for Tristram's Storm-Petrel (TRSP) in the Pacific Seabird Database, but limited data are available from Tern Island between 1995 and 2002 (Fig. 2.25) and a graduate student (Greg McClelland) recently collected data on Laysan Island. We only have access to the data from Tern Island and these data come from a mixture of natural burrows and nest boxes. To our knowledge, breeding success is only intermittently monitored. Like Bonin Petrel and Bulwer's Petrel, over 50% of the known population of TRSP breed within the US Pacific Islands.

Prospective power of MIC data and assessment of current methods

TRSP incubate their eggs between 40 and 45 days. We did not have access to data from Laysan and we did not think there was enough information from Tern to accurately estimate prospective power. We suspect there are more than 32 nesting pairs on Tern Island and that observers mainly monitor nest boxes. However, if this is truly a census, there are still not enough data to address prospective power.

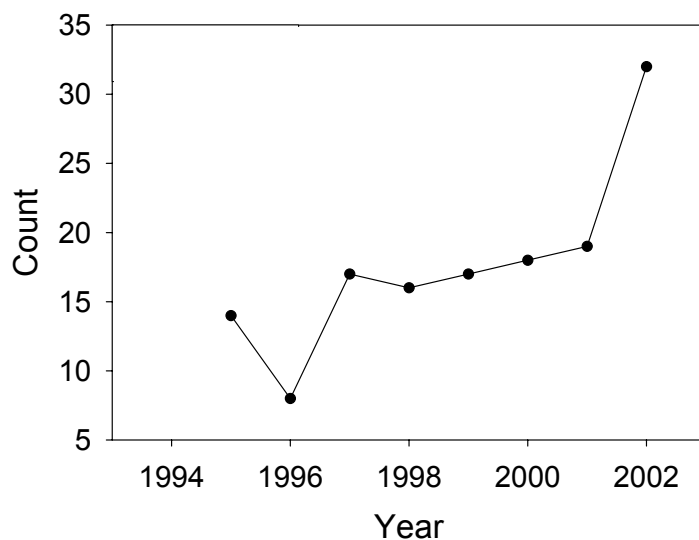


Figure 2.25. Count data for TRSP on Tern Island between 1995 and 2002.

Precision of reproductive data

Again, we did not think there was enough information to accurately estimate the precision of current sampling methods.

Summary and recommendations

Count data. On Tern Island, we suspect the limiting factor for detecting trends is likely population size, not how the population is monitored (as long as counts are generally accurate). We do not think that improving sampling designs for TRSP will improve power. If managers are interested in population trends, we recommend that current methods are continued with the understanding that data analysis will be qualitative. In other words, the interpretation of count data will rely on non-statistical measures. As such, we do not think that banding individuals on Tern Island is a worthwhile task. Sample sizes are definitely too low for estimating survival with banded individuals and TRSP are highly susceptible to disturbance.

Populations on Nihoa, Laysan Island, and Pearl and Hermes Reef are reported to be in the thousands (USFWS 2005). We suggest using the estimators for a simple random sample if these populations are monitored (see Recommendation 4, Section III).

Reproductive data. We suggest using the estimator for reproductive success that is presented in Section IV.

BLUE-GRAY NODDY (*PROCELSTERNA CERULEA*)

Data availability

No data are available within the planning area for Blue-Gray Noddies (BGNO). BGNO are known to have large colonies on Necker and Nihoa Islands, but small colonies exist on other islands, including La Perouse at French Frigate Shoals. Counts have been conducted by boat at La Perouse; however these data are intermittent and were not made available for this manual. BGNO are known to nest within rocky cavities on exposed cliffs and sea stacks. They will sometimes nest in soil (USFWS 2005). Average incubation length is 35 days. Individuals breed between March and June on French Frigate Shoals, but no data are available for other locations (USFWS 2005).

Recommendations

Little is known about BGNO. Base-line studies and surveys will be necessary to determine how best to monitor this species. However, because BGNO are one of only two neuston-feeding terns in this assessment, we included them in this document. Until the monitoring program goals for other species are accomplished, we suggest considering BGNO a low priority species for monitoring.

We suspect that alternative monitoring methods (i.e., a different monitoring framework) will be necessary for BGNO. Because the only large colonies are on islands that are difficult to reach and because individual nest-sites will be difficult to monitor, sampling the number of breeders or sampling reproductive success will be difficult. In particular, in response to Goal 2, to evaluate the species as an indicator of neuston food resources, we suggest corticosterone analysis (Section V). On each island, a small sample (approx. 10 birds) may yield insight as to the availability of food. However, the method will have to be calibrated and little is known regarding the timing of breeding at different locations. In other words, we may not know when to sample breeders.

Another possibility is to count individuals at sea, near known breeding colonies. BGNO are considered inshore feeders (Rauzon et al. 1984) and are rarely encountered far at sea during breeding (USFWS 2005). At sea surveys could be conducted from boats that follow transect lines. Distance sampling is one possible method (e.g., Buckland et al. 2001). Recently, much

work has been done with at sea surveys and generalized additive models (GAMs). The methodology, as generally applied to seabirds, is described in Clarke et al. (2003). In short, GAMs are a semi-parametric regression. The method is usually applied to situations where the number of individuals counted is small (i.e., small sample sizes), as is often encountered with at-sea surveys. The regression line is determined by non-parametric smoothing (versus the typical sum-of-squares methodology) and the error distribution (for estimating standard errors and confidence intervals) are parametrically estimated. The results of Clarke et al. (2003) suggest that at-sea surveys may correlate with the abundance of breeding birds on cliffs. However, we recommend the method be calibrated by comparing actual breeding counts with the at-sea surveys. After the correlation between at-sea surveys and breeding populations has been quantified, the method can be considered for wider application.

GRAY-BACKED TERN (*STERNA LUNATA*)

Data availability

Tern Island. We used 24 years of data (1980-2003) where Gray-Backed Terns (GRAT) are counted on the entire island (Fig. 2.26). In addition, six years of reproductive data are available between 1980 and 1996.

Midway Atoll. Two years of count data are available for Midway Atoll (1995-1996). This was not enough data to be used for analyses.

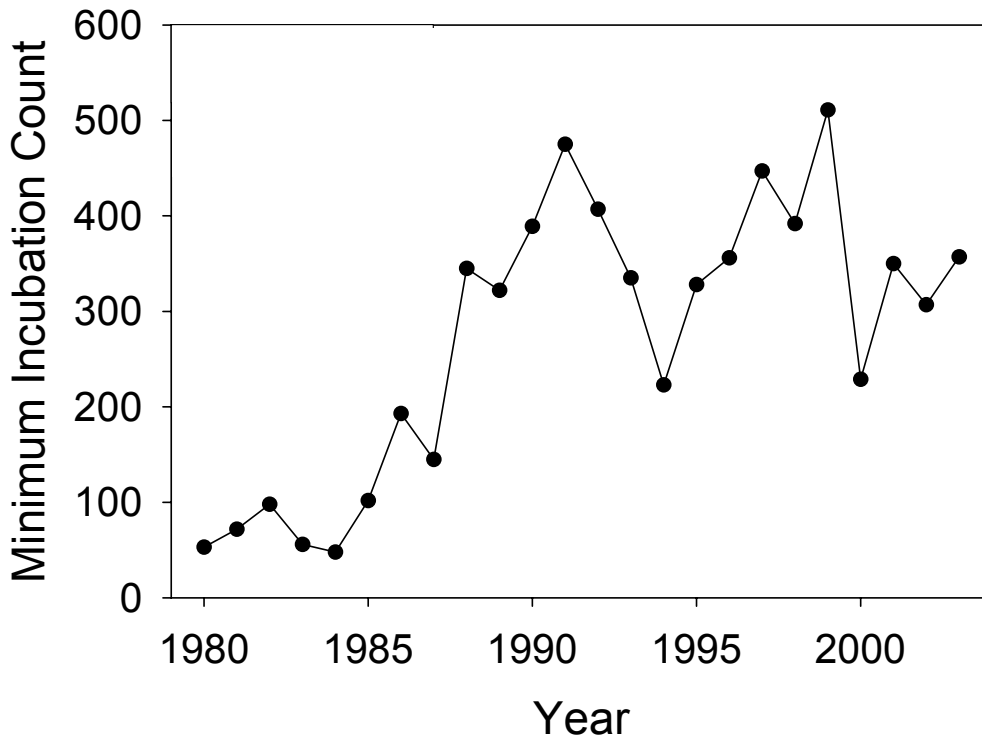


Figure 2.26. Mean incubation counts for GRAT on Tern Island between 1980 and 2003.

Prospective power of MIC data

We determined that variance increased with population size on Tern Island (Fig. 2.26) and log transformed count data to stabilize errors. We did not detect serial autocorrelation in the data series. With current methods, data would have to be collected for 15 years on Tern Island (Table 2.10). Within a 10 year period, only declines greater than 80% are detectable on Tern Island (Table 2.11).

Table 2.10. Years of observations required to detect trends in GRAT mean incubation counts.

Standard	Tern Island (SD=0.3734) ^a
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.05$; power = 0.9 ^b	15
50% decline over 10 years; 6.7% annual decline; $\alpha = 0.1$; power = 0.9 ^c	13

^a Standard deviation on a log scale.

^b Recommendations of Hatch (2003).

^c Proposed standards for the Alaska Maritime National Wildlife Refuge (G.V. Byrd, pers. comm.).

Table 2.11. Change in GRAT mean incubation count detectable over 10 years.

Standard	Detectable change over 10 years
	Tern Island
$\alpha = 0.05$; power = 0.9	80%
$\alpha = 0.1$; power = 0.9	75%

Assessment of current methods

On Tern Island, MIC counts for GRAT are highly variable and prospective power is low. Low power is likely due to low sample size combined with very high variability. We did not have data from other locations and do not know if GRAT counts are highly variable elsewhere. Population sizes are likely larger on Lisianski, Nihoa, and Laysan (USFWS 2005). Average length of incubation is 30 days.

Alternatives

Data from Tern Island suggest GRAT are only moderately asynchronous in breeding (Fig. 2.27). In other words, breeding is generally peaked and it is unlikely that either the use of alternative metrics as described in other sections or the mark-recapture ratio estimator (as described for BLNO and in Recommendation 8 of Section II) will provide large gains in power. Although the mark-recapture estimator will provide estimates that better approximate the true breeding population, we suggest that the application of this experimental estimator is better suited to other species, such as Black Noddy, that are highly asynchronous, or Red-tailed Tropicbird, where banding data are currently collected. Also, GRAT are highly susceptible to disturbance, so banding may not be a feasible option.

Precision of reproductive data

Although some reproductive data are in the Pacific Seabird Monitoring Database, there is

no plot level data that allows examining levels of precision. We suggest using the estimator provided in Section III.

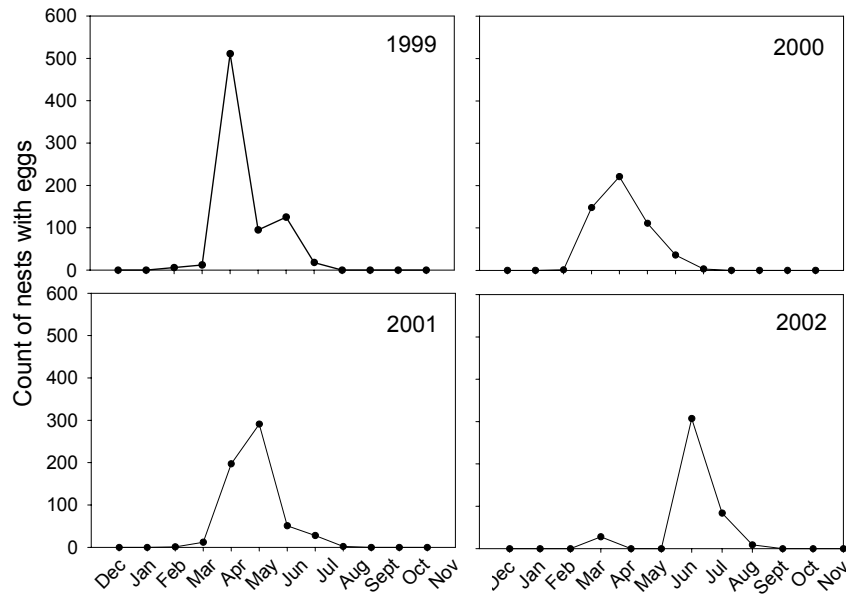


Figure 2.27. Mean incubation counts of GRAT by month on Tern Island between 1999 and 2002.

Summary and recommendations

Count data. We conclude that the ability to detect trends in GRAT breeding populations is below standard at Tern Island. However, this is one of only two neuston feeding terns in the manual and GRAT will be much easier to monitor than the other neuston feeding tern, Blue-Gray Noddy. If managers want to monitor a neuston feeding tern, GRAT is the best target species, even if counts will have low power. However, if keeping track of food resources is of more interest than the species itself, we recommend focusing monitoring efforts on reproductive success (see below).

Reproductive data. We suggest applying the estimators for reproductive success in Section IV. Although the statistical power of population counts will likely be low, we see no reason why precise estimates of breeding success cannot be quantified. One of the main reasons that representatives from each foraging guild were chosen, was to build in some capabilities to assess causal relationships if a decline is noticed, e.g., if all the tuna birds begin experiencing population declines or reduced breeding success while neuston feeders did not, biologists can then generate a hypothesis related to marine resources.

SECTION III: SAMPLING DESIGNS FOR ESTIMATING BREEDING COUNTS

What follows is a decision tree with recommended protocols and estimators. We intend this decision tree to help guide managers, not as a complete description for how to sample in any particular situation. These protocols will have to be amended slightly for particular situations, but they will form the basis for any sampling design. We encourage biologists and managers to consult with statisticians and quantitative biologists about the application of these techniques. The authors of the report will be happy to consult with seabird managers in the development of appropriate sampling designs or in implementing pilot studies. Contact Michelle Reynolds at USGS Pacific Island Ecosystems Research Center: michelle_reynolds@usgs.gov. We focus on quantifying the number of breeding birds within years. At the request of the USFWS, we tried to avoid suggesting mark-recapture estimators unless absolutely necessary. We realize that analyzing data from marked individuals (e.g., Lebreton et al. 1992) will yield information much more detailed and useful than simple counts of breeding birds. For example, mark-recapture estimators allow the estimation of total population size (i.e., not just the number of breeders within years), the number and probability of individuals that skip breeding seasons, and provide information that could be used to parameterize population models (e.g., survival and dispersal probabilities). Mark-recapture estimators will also allow for incomplete and variable detectability. We acknowledge that the logistics greatly limit monitoring of remote sites, that USFWS has limited resources, and would like to avoid handling birds if possible. More robust monitoring methods might be considered in the future as funding and personnel allow.

The framework of our decision tree reflects what we believe to be the main issues for counting breeding birds within the planning area. We think the main issues associated with counting breeding birds are these:

- 1) Can all active nests in plots be located (i.e., is detectability an issue)? If so, what level of search effort is required? Are mark-recapture estimators (for nests) required?
- 2) If all active nests in plots can be located, can observers also count nests across the

entire island or are observers logistically limited to counting nests on plots? Are the plots a random sample of the colony?

3) Is the species a synchronous breeder (Section III, Table 3.1)? In other words, do we expect the MIC to be strongly correlated with the true number of breeders?

Based upon the answers to these general questions, we provide a recommended course of action. Where practical, we provide estimators in spreadsheets attached to this report. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed. Finally, we provide two examples of how to use the decision tree to select a sampling design.

Table 3.1. Breeding synchrony for the species considered in this manual. Note that we do not have a robust metric of synchrony to classify species as exhibiting high or low synchrony. All data are from Tern Island. We recommend this table be considered cautiously, as levels of synchrony may differ by location and year. For example, anecdotal data suggest that Red-footed Boobies are less synchronous on Palmyra Atoll than on Tern Island.

Group	Species	Seasonality	Synchrony
Pelecaniformes	Red-footed Booby	Aseasonal	Low
	Red-tailed Tropicbird	Aseasonal	Low
	Least Frigatebird	Aseasonal	Low ??
	Great Frigatebird	Aseasonal	Low
Tuna birds	Sooty Tern	Aseasonal	Low
	Black Noddy	Aseasonal	Low
	Wedge-tailed Shearwater	Seasonal	High
	Christmas Shearwater	Seasonal	High
Nocturnal petrels	Bulwer's Petrel	Seasonal	High
	Bonin Petrel	Seasonal	High
	Tristram's Storm-petrel	Seasonal	High
Neuston terns	Blue-gray Noddy	Aseasonal	High ??
	Gray-backed tern	Aseasonal	Low ??

Decision Tree for Estimating the Number of Breeding Seabirds

Question 1: Are breeders synchronous (See Section III, Table 3.1)?

YES.....go to Question 2

NO.....go to Question 5

Question 2: What methods are necessary to ensure complete detectability (i.e., a census) on study plots? Before answering this question, go to Recommendation 1 and estimate detectability.

YES, all active nests are easily counted on study plots...go to Question 3

YES, but intensive searching is required to guarantee a census.....go to Question 4

NO.....go to Recommendation 2

Question 3: Can active nests be counted on the entire island?

YES.....go to Recommendation 3

NO.....go to Recommendation 4

Question 4: Can active nests be counted on the entire island?

YES.....go to Recommendation 5

NO.....go to Recommendation 6

Question 5: What methods are necessary to ensure complete detectability (i.e., a census) on study plots? Before answering this question, go to Recommendation 1 and estimate detectability.

YES, all active nests are easily counted on study plots...go to Question 6

YES, but intensive searching is required to guarantee a census.....go to Recommendation 7

NO.....go to Recommendation 8

Question 6: Can active nests be counted on the entire island?

YES.....go to Recommendation 9

NO.....go to Recommendation 10

RECOMMENDATIONS FOR ESTIMATING THE NUMBER OF BREEDING SEABIRDS

Recommendation 1

In many cases observers can completely count (i.e., census) active nests on study plots. We strongly suggest that the ability to completely count the number of active nests be tested. Many of the sampling methods we recommend depend upon having a complete count or census on study plots. If this assumption is not valid, then resulting estimates will not be valid.

We suggest using a simple Lincoln-Petersen estimator (e.g., Lancia et al. 1996) with an adjustment for small sample size (Chapman 1951). The recommended protocol is:

- 1) On study plots, have one observer mark active nests with an unobtrusive marker (such as washers or small flags that will not be visible to the next observer). On each plot, this value is n_1 .
- 2) Have another observer recount nests on the study plot. The total number of nests counted and the number marked nests that the second observer identified are recorded. The total number of nests counted equals n_2 . The number of previously marked nests the second observer identifies is equal to m_2 .
- 3) For each plot, record the values of n_1 , n_2 , and m_2 in the spreadsheet we provide for estimating plot level detection (see “Estimator 1.xls”). This spreadsheet will calculate detectability (i.e., probability an individual nest is detected).

We recommend that if detectability is not consistently greater than 0.95, that the assumption of complete detection of nests on study plots is invalid. We recommend that detectability be estimated on plots for each species and that this assumption be tested periodically (e.g., once every few years) to make certain the assumption holds.

If complete counts on plots are not possible (i.e., if detectability is below 0.95), alternative estimators are needed.

Recommendation 2

If nesting birds cannot be reliably counted on study plots, then we recommend that a mark-recapture or distance sampling estimator be used to estimate the probability of detection for nests or breeders. A number of different options are possible. For example, the detectability estimator in Recommendation 1 can be used in conjunction with the estimators provided in Recommendation 5 or 6 to adjust for detectability. We propose that this approach may improve statistical power to detect trends for asynchronously nesting species. However, we suspect that other designs may be more efficient in some situations. For example, on Palmyra, distance sampling will be a useful framework. If a complete count is not possible on study plots, a more complicated approach will be needed. Refuge staff should consult the authors of this report or a quantitative biologist.

Recommendation 3

If birds are synchronous breeders and the entire population of attended nests on an island can be counted without error, then a true census exists. We suggest not assuming a census exists unless the detectability is consistently over 95%. In other words, we suggest estimating detectability with other methods, before making this assumption. Use the Excel spreadsheet "Estimator 1.xls" to test this assumption. Scenarios where a census may apply are for synchronous breeders, such as the shearwaters, petrels, and storm-petrels on small islands, such as Tern Island or possibly for small colonies on other islands.

Recommendation 4

For synchronous breeders, if active nests can be completely counted on study plots (i.e., we can realistically assume a census on study plots), but not across the entire island, this plot-based method described below is the general protocol to use. For example, this is likely the appropriate estimator for the shearwaters, petrels, and storm-petrels, either on large islands (such as Laysan Island or Johnston Atoll), or when observers want to minimize disturbance by sampling only a portion of the population. This is the general protocol:

- 1) Plot size must be determined. A general recommendation is for plots to be large

enough to contain 20 nests (See Appendix D for explanation). This number is flexible however depending on biology of the species and the layout and topography of the breeding colony.

2) Observers must map the colony boundaries and calculate the total number of possible plots (N ; this is equal to colony area divided by plot area), then randomly select n plots from the total number of plots N .

3) On each plot, n , count the number of attended nests or burrows. We highly recommend using a burrowscope if nest contents are not visible. Managers should choose an operational definition for an “active burrow” (e.g., courting pair, incubating pair, etc.) that meet the needs of the specific monitoring program.

4) An Excel spreadsheet is attached that contains the appropriate equations and citations for a simple random sample abundance estimator. This spreadsheet will estimate abundance and 95% confidence intervals. This spreadsheet is titled: “Estimator 2.xls”.

This estimator is based upon simple random sampling. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed. For more information see: Scheaffer, R. L., W. Mendenhall, and L. Ott. 1996. Elementary survey sampling. Fifth edition. Wadsworth Publishing Company, Inc. Belmont, CA. The actual equations are given on page 91.

Recommendation 5

If complete counts of active nests on study plots (i.e., a census of study plots) are only possible with “intensive searching”, but observers can conduct less intensive count nests on the entire island, then the estimator described below may be useful. The general idea is that two counts are made on plots. The first is a “rapid” count and the second is an “intensive” count. If the “intensive” count results in complete detection, then this information can be used to estimate detectability and to adjust the island-level count.

For this method to work, the “intensive” count must result in a plot-level census. Intensive search methods should be used with the Lincoln-Petersen estimator in Recommendation 1 to guarantee this assumption is met. The general protocol is this:

- 1) Designate detectability plots. For consistency, we recommend making these plots the same size as reproductive plots.
- 2) Using the same methods that are used for MICs, count the number of nests on the plot (i.e., this is the “rapid count”).
- 3) A different person should search the plot intensively and count the number of nests on that plot (i.e., this is the “intensive” count).
- 4) This leads to three types of information collected during each MIC: (a) the island-wide “rapid” count (i.e., the MIC), (b) a “rapid” count on each plot, and (c) an “intensive” count on each plot.
- 5) The same observer should conduct the MIC and the plot level “rapid count”. A different observer should conduct the plot level “intensive” count.
- 6) An Excel spreadsheet is attached that contains the appropriate equations and citations for adjusting MICs for detectability when the entire island is counted. This spreadsheet will estimate detection probability, abundance, and 95% confidence intervals. This file is: “Estimator 3.xls”.

For more information, see: Thompson, S. K. 2002. Sampling. John Wiley and Sons, Inc., New York, NY, USA.

Recommendation 6

When a complete count or census of active nests is only possible on study plots when observers conduct intensive searches, but the entire island cannot be sampled, most authors suggest “double sampling” (e.g., Bart and Earnst 2002; Thompson 2002). Double sampling refers to designs in which initially a sample is selected for “auxiliary” information (such as a “rapid” count of nests) and then a subsample is selected for obtaining both auxiliary and detailed information (such as conducting both “rapid” and “intensive” counts). For example, we can conduct rapid counts on many plots and then conduct both “rapid” and “intensive” counts on a subsample.

As in the previous recommendation, for this method to working the “intensive”

count must result in a plot-level census. Intensive search methods should be used with the Lincoln-Petersen estimator in Recommendation 1 to guarantee this assumption is met.

The input and protocol is similar to that of Recommendation 5 (where rapid and intensive counts are used, but the entire island is sampled). The exception is that the observer must record the number of rapid count plots (n'), intensive count plots (n), and sum of nests from all rapid count plots. The protocol is as follows:

- 1) Designate “rapid count” plots. For consistency, we recommend making these plots the same size as reproductive plots. Using the same methods that are used for MICs, conduct a rapid count of the number of nests on the plot.
- 2) Designate a subsample of rapid count plots to be used as both rapid count plots and intensively searched plots. First count the nests on the plot. Then have a different person intensively search the plot.
- 3) The same observer should conduct the MIC and the plot level rapid count. A different observer should conduct the plot level intensive search.
- 4) An Excel spreadsheet is attached that contains the appropriate equations and citations for adjusting MICs for detectability when the entire island is not counted. This spreadsheet will estimate detection probability, abundance, and 95% confidence intervals. This file is: “Estimator 4.xls”. The variables involved in this calculation are n' = the number of rapid count plots, n = a subsample of n' ; the number of plots where both rapid counts and intensive searches occur, x_i = the number of nests counted within a rapid count on plot i , y_i = the number of nests counted when intensively searching on plot i , and N = the total number of possible plots (of any kind).
- 5) If detectability is over 95%, we suggest assuming that a complete census is valid. In other words, if detectability is consistently over 95%, we would not recommend estimating detectability for every MIC.

For more information, see: Thompson, S. K. 2002. Sampling. John Wiley and Sons, Inc., New York, NY, USA. The actual formula is on page 159 of Thompson

(1992). *IMPORTANT: On most islands, the extent of colonies is limited. Because most islands have a small area and a limited number of possible plots, observers may choose to have equal numbers of plots for “rapid” and “intensive” counts. If so, do not use this estimator, as it does not yield valid estimates of variance when $n' = n$. Instead, use the estimator for a simple random sample when the entire island cannot be sampled (“Estimator 2.xls”).*

Recommendation 7

Within this scenario, birds are not synchronous breeders and we must correct for asynchrony. However, observers must also correct island-wide counts for detectability. While we have developed an estimator for this scenario (based upon principles given by Cochran [1977] and Thompson [2002]), we hesitate to recommend this estimator, as it is very complicated. The issue is that to correct for asynchrony, observers will have to monitor plots with marked individuals. To correct for detectability, separate detection plots will have to be monitored as we are concerned that using the same plots for banding and estimating detection will not lead to unbiased estimates of detectability. In short, we suspect that if banding plots are intensively monitored, that observers will have pre-existing knowledge of where to find nests. Hence, if these plots are used to estimate detectability, estimates of detectability will always be 100%, when true detectability will be lower. Hence, this method would require having two separate types of plots. We doubt that logistical constraints will permit such a design and, if logistical constraints are not limiting, then a specially tailored mark-recapture design will be more appropriate as it will yield more reliable and useful information.

Hence, we make the following suggestion: Use the design in Recommendation 10 for adjusting MICs for asynchrony when the entire island is not sampled (“Estimator 6.xls”). Because we can assume that banding plots are intensively searched, this will be a valid estimator. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed.

Recommendation 8

If birds are asynchronous breeders, such that the number of active nests is not a reliable measure of the number nesting pairs, then we recommend that mark-recapture

methods be used. By marking individual birds, we can account for both asynchrony and detectability. A sampling design will have to be tailored for the individual situation.

Recommendation 9

If birds are asynchronous breeders and can be counted on the entire island without error, the following sampling design is recommended. The general idea is to use information from currently monitored study plots to determine how the number of total breeders relates to mean incubation counts. In effect, we compute a conversion factor on the reproductive plots and apply this to the entire population.

The theory comes from Cochran (1977) and Thompson (1992). Assume the island consists of N possible plots, of which n are sampled. Also assume that the total population count is M_{total} ; this is the total number of nests (nests with eggs) counted on the island. The plot level counts are m_i for plots 1- n (i.e., plot 1 is m_1 , plot 2 is m_2 , etc.). On each plot, mark all birds. The number of unique pairs on each plot equals y_i for plots 1- i . The ratio of interest is:

Eq. 3.1)

$$r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n M_i}$$

The number of breeders is estimated as:

Eq. 3.2)

$$\# \text{ pairs} = rM_{total}$$

This equation is the ratio multiplied by the total count. For example, assume there is only one plot. If observers counted 100 egg nests on this plot, then $M_1=100$. By keeping track of banded birds, we find there were many instances of renesting. Assume there were only 50 unique pairs on this plot, the ratio is equal to $50/100 = 0.5$. If the total count is 1,000, then the estimated number of breeders is $r*M_{total} = 0.5*1000 = 500$. This method simply adjusts the total count of nests for the number of unique breeders on plots. The

general protocol is this:

- 1) Observers conduct Mean incubation counts (MICs). The counts from each month are totaled (i.e., sum all MICs within a year). This is entered as M_{total} on the attached spreadsheet titled “Estimator-Adjust MICs for asynchrony with banding data.xls”. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed.
- 2) On plots, observers keep track of the total number of pairs within the plot during each MIC. Observers must use bands to keep track of the total number of unique pairs. The number of unique pairs for the entire year is entered into the column “Number unique pairs (y_i)”.
- 3) Hence, at the end of each breeding season, observers have three types of data: (a) sum of MICs for the entire island, (b) sum of MICs specific to each plot, and (c) count of the number of unique breeders on each plot.
- 4) We have attached an excel spreadsheet that details how to adjust MICs for asynchrony when the entire island is counted. This spreadsheet can be used to calculating the ratio, the # nests, the variance of # nests, and the 95% CI. This spreadsheet is titled: “Estimator 5.xls”.

In practice, bands will have to be used to determine the number of unique pairs, not the number of unique bands. If individuals are monitored on plots, the total number of unique breeding pairs can easily be determined.

We suspect this will be the best estimator for many asynchronous species, such as Red-tailed Tropicbirds, Red-footed Boobies, and Black Noddies (See Section III, Table 3.1). However, we realize that this estimator cannot be applied to all asynchronous species simultaneously. We suggest that this limited mark-recapture estimator be applied to one or two species.

Recommendation 10

If birds are asynchronous breeders and birds cannot be counted on the entire

island for logistical reasons (such as island size or concerns regarding disturbance), we suggest the following sampling design. *This estimator is very similar to the one presented in Recommendation 9, but observers do not count all nests on the island, only nests on plots.* As such, the estimator for population total is the same, but variance estimators differ. The general idea is to use information from currently monitored study plots to determine how the number of total breeders relates to mean incubation counts. Thus, we compute a conversion factor on the reproductive plots and apply this to the entire population. See Recommendation 9 (above) and Appendix C for more information. The appropriate estimators for adjusting MICs for asynchrony when the entire island is not counted are provided in the spreadsheet titled: “Estimator 6.xls”. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed.

Example 3.1: Wedge-tailed Shearwaters on Tern Island.

Decision tree for how to estimate the number of breeding birds

Question 1: Are breeders synchronous (See Section III, Table 3.1)?

YES.....go to Question 2

NO.....go to Question 5

Answer: YES. *Wedge-tailed Shearwaters are synchronous breeders (See Section III, Table 1). Go to Question 2.*

Question 2: What methods are necessary to ensure complete detectability (i.e., a census) on study plots? Before answering this question, go to Recommendation 1 and estimate detectability.

YES, all nests are easily counted on study plots...go to Question 3

YES, but intensive searching is required to

guarantee a census.....go to Question 4

NO.....go to Recommendation 2

Answer: YES, all nests are easily counted on study plots. *First estimate detectability. We expect that detectability will be high and that burrows can easily be counted on plots with the help of a burrowscope. Go to Question 3.*

Question 3: Can nests be counted on the entire island?

YES.....go to Recommendation 3

NO.....go to Recommendation 4

Answer: YES. *We expect that on Tern Island, observers can count nests on the entire island. If detectability is high and the entire island can be sampled with the same methods, then we concluded that a census is reasonable. See Recommendation 3.*

Example 2: Wedge-tailed Shearwaters on Laysan Island and Midway Atoll.

Decision tree for how to estimate the number of breeding birds

Question 1: Are breeders synchronous (See Section III, Table 3.1)?

YES.....go to Question 2

NO.....go to Question 5

Answer: YES. *Wedge-tailed Shearwaters are synchronous breeders (See Section III, Table 1). Go to Question 2.*

Question 2: What methods are necessary to ensure complete detectability of nests (i.e., a census) on study plots? Before answering this question, go to Recommendation 1 and estimate detectability.

YES, all active nests are easily counted on study plots...go to Question 3

YES, but intensive searching is required to

guarantee a census.....go to Question 4

NO.....go to Recommendation 2

Answer: YES, all active nests are easily counted on study plots. *First estimate detectability. We expect that detectability will be high and that burrows can easily be counted on plots with the help of a burrowscope. Go to Question 3.*

Question 3: Can active nests be counted on the entire island?

YES.....go to Recommendation 3

NO.....go to Recommendation 4

Answer: NO. We expect that on larger islands, like Laysan Island and Midway Atoll, that observers cannot count nests on the entire island. In such cases, we suggest using the Simple Random Sample estimator provided in Recommendation 4. Observers monitor a random sample of plots.

SECTION IV: SAMPLING DESIGNS FOR ESTIMATING REPRODUCTIVE SUCCESS

On most islands, reproductive success is monitored on plots, assuming that plots are representative of breeding success on the island as a whole. Because plot-based monitoring is more efficient than monitoring all nests, or a random sample of all nests, we support plot-based studies of reproductive success. This sort of sample is referred to as a “cluster sample,” because each sample contains a cluster of nests. Thus, we are concerned with randomly sampled plots, not nests. We provide estimators for reproductive success in spreadsheets attached to this report. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed.

PLOT SELECTION ISSUES

The importance of randomization

How plots are selected is a critical issue. If plots do not represent the population, then our estimates of reproductive success are not representative of the population. The only way to guarantee that plots represent the population is through random plot selection. We realize that there are logistical constraints associated with plot selection and that observers may prefer to monitor existing plots. Monitoring existing plots is acceptable, if existing plots are representative. We suggest that managers critically examine this issue.

Resampling existing plots

A fundamental question if plots are to be sampled repetitively is: whether the same plots be sampled each year or should new plots be sampled each year? Again, the degree to which plots represent the entire population is important. By selecting new plots each year, we help ensure that plots represent the entire population. However, at the same time, we increase the level of variation in our estimates of breeding success (or any other metric of reproduction). When examining changes in reproductive success over time, sources of variation include annual variation (i.e., the yearly change in reproductive success) and variation among plots within each year (each plot has a different level of reproductive success). If we choose a new sample of plots, then we increase the variation among plots. As we add more variation to our estimates of reproductive success, it becomes more difficult to estimate trends. Hence, the most precise estimates of trends are derived from sampling designs where the initial plot selection is random

and then the same plots are sampled following years (Thompson et al. 1998, page 177). Sampling the same plots each year also introduces patterns of autocorrelation, but this can be accommodated with repeated measure designs (e.g., Usher 1991).

We recommend resampling existing plots. If there is concern that existing plots are not representative, we recommend sampling some new (randomly selected) plots and comparing reproductive success between new and old plots.

Changes in the distribution of nests (empty plots)

Colony boundaries may shift and change. If plots are no longer occupied by birds because the colony has shifted in distribution, select some new plots. By randomly selecting new plots (in appropriate habitat), we ensure that our sample represents the population and will actually increase precision in our estimates. This is because we will no longer have plots with zero nests.

SAMPLING DESIGN

- 1) The boundaries of the colony need to be mapped and colony area should be calculated. This can be accomplished within GIS. We recommend that general colony boundaries be verified on a yearly or bi-yearly basis.
- 2) Within the colony, plots should be randomly located. There are two methods for random plot assignment, depending upon whether or not the sampling frame is well defined (Fig. 2.21). If the sampling frame is well defined, then a sample of possible plots is selected. If the sampling frame is not well defined (i.e., if there is an “open frame” within a defined boundary), then random x and y coordinates are chosen, rather than plots. In theory, statistical estimators rely on the random selection of plots, not x-y coordinates. However, the biases are negligible in practice. Especially in real situations, where colony boundaries may not conform to plot shapes, the random selection of x-y coordinates is preferred. Plots must be of equal *area*, but do not have to be circles or squares. For species that nest along narrow strips of habitat, do not hesitate to create irregular plots (as long as the additional plots are of the same total *area*).
- 3) If nests are strongly clustered, we suggest recording nest locations with a GPS and then defining clusters. Instead of selecting random plot locations, center plots on clusters of nests and then randomly select nest clusters.

4) The required plot sizes and the number of plots will depend upon the variation between plots, the total number of possible plots, and the desired level of precision. As the variation among plots increases, more plots will have to be sampled. For a given level of variation, if the number of possible plots is large, more plots will also have to be sampled to ensure the sample is representative. As a start, we recommend having a minimum of eight plots that are big enough so that there are more than 25 active nests in each plot. Even when the variation between plots is large, precise estimates are possible with small numbers of plots (<10) when there are more than 25 active nests in each plot (See Appendix D). We suggest estimating precision each year using the estimator we provide in the attached spreadsheet (“Estimator 7.xls”). This estimation can be accomplished by using the plot level data to calculate the 95% bound (the spreadsheet will do this for you). We suggest trying to achieve a 95% bound of either 0.05 or 0.1. In other words, with an estimate of breeding success of 0.5, if the 95% bound is 0.1, we are confident that the true value is somewhere between 0.5 ± 0.1 (i.e., 0.4 to 0.6). This level of precision (95% bound = 0.05 or 0.1) will likely be sufficient for most applications of reproductive data.

5) Once plots are randomly selected, we recommend resampling the same plots each year. Some authors will suggest sampling different plots each year, but resampling the same plots will yield the greatest power to detect trends (Thompson et al. 1998). The key issue is that plots must be representative of the breeding population. The initial random selection of plots helps ensure that plots are representative. We suggest that observers regularly question if plots are representative. If there is doubt that sampled plots are representative of unsampled plots, we suggest adding plots and comparing new plots to old plots. A special case is due to observer effects. For example, if observers are causing low breeding success on plots, then plots may not provide an unbiased measure of breeding success on the island. However, adding plots will not be able to address if plots are representative, unless different methods of observation can be compared. Focused studies designed to compare different levels of disturbance will be necessary to evaluate observer effects.

6) We strongly suggest that all plot level data are recorded and archived. We recommend saving a copy of the attached completed spreadsheets for each breeding season, both in hard copy and on computer media. If the plot level data are archived, this data can be applied to alternative estimators or be used to assess sampling design in the future.

ESTIMATOR FOR REPRODUCTIVE SUCCESS

The estimator currently used by the USFWS (1980) is from Kish (1965). This estimator is for “cluster sampling”. Cluster sampling occurs when objects of interest are more effectively or efficiently sampled in groups (i.e., clusters), than singly. For example, it is much easier to select a random sample of plots that contain clusters of nests than to randomly sample individual nests. In these estimators, the number of clusters is equal to the number of plots. The estimator from Kish (1965) is:

Eq. 4.1)

$$\hat{P} = \frac{\sum a_i}{\sum m_i}, \quad \text{E.g.: } \textit{breeding_success} = \frac{\sum \textit{fledged}}{\sum \textit{eggs}}$$

where \hat{P} is the estimate of breeding success, a_i is number of individuals fledged on plot i , and m_i = number of individual eggs on plot i . This is the correct estimator for breeding success. However, the variance estimator that was in use was incorrectly copied from Kish.

The reference on this estimator is Kish (1965). The most simplified form of this estimator (from Kish) is:

Eq 4.2)

$$\text{var}(\hat{P}) = \left(\frac{1-f}{m^2} \right) \left(\frac{n}{n-1} \right) \left(\frac{1}{\sum m_i} \right) \left(\sum a_i^2 + \hat{P}^2 \sum m_i^2 - 2\hat{P} \sum m_i a_i \right)$$

Where, N = total number of clusters (i.e., number of possible plots), n = clusters sampled (i.e., number of sampled plots), and:

$$f = \frac{n}{N}$$

The estimator currently used does not account for the proportion of the population sampled, thus there is no benefit for sampling lots of clusters and no penalty for sampling few. However, as the proportion of the possible clusters sampled increases, variance will approach zero. We recommend using more commonly used estimators available in Cochran (1977), Raj (1968), and

Mendenhall et al. (1971), and Scheaffer et al. (1996). We think the best variance estimator is from Scheaffer et al. (1996, Elementary survey sampling, page 137):

Eq. 4.3)

$$\text{var}(\hat{P}) = \left(\frac{N-n}{Nn\bar{m}^2} \right) \left(\frac{\sum_{i=1}^n (a_i - \hat{P}m_i)^2}{n-1} \right).$$

Where, N = total number of clusters (i.e., number of possible plots), n = clusters sampled (i.e., number of sampled plots), a_i = number of individuals fledged on plot i , m_i = number of individual eggs on plot i , p_i = the proportion of individuals fledged/eggs on plot i , and

$\bar{m} = \frac{1}{n} \left(\sum_{i=1}^n m_i \right)$ = the average number of individuals per plot. This estimator actually calls for

\bar{M} instead of \bar{m} ; where $\bar{M} = \frac{1}{N} \left(\sum_{i=1}^N m_i \right)$, the average number of individuals on all plots in the

population. Scheaffer et al. (1996) state that \bar{m} may be used to approximate \bar{M} .

We have attached an Excel spreadsheet that will calculate the estimator for reproductive success: “Estimator 7.xls”.

The variance estimator differs from that of Cochran (1977) in that Cochran does not penalize for the proportion of clusters sampled in the same fashion that Scheaffer et al. (1996) does. Cochran’s estimator is:

Eq. 4.4)

$$\text{var}(\hat{P}) = \left(\frac{N-n}{Nn\bar{m}^2} \right) \left(\frac{\sum_{i=1}^n m_i (p_i - \hat{P})^2}{N-1} \right)$$

This estimator is analytically identical to the estimator of Scheaffer et al. (1996) except that the variance estimator is divided by $N-1$ instead of $n-1$. Note that although the numerators of these variance estimators look different, they are analytically equivalent (i.e., inserting the same values yields the same results). We use the estimator of Scheaffer et al. (1996), because dividing by number of sampled plots minus 1 (i.e., $n-1$) better indicates how decreasing the number of sampled plots increases variation. When dividing by the total number of plots minus 1 (i.e., $N-1$) there is no penalty for sampling few plots.

SECTION V: GENERAL MONITORING RECOMMENDATIONS

We have generated two lists of monitoring recommendations. The first list identifies five specific monitoring priorities, ordered from most important to least important. Second, we make a series of general recommendations that are widely applicable to a variety of decisions that managers may make when designing or implementing seabird monitoring programs. We hope that these recommendations will provide a useful framework for modifying existing monitoring programs and designing new ones.

In preparing this document, we identified several topics where we feel that additional research can contribute to the goals of seabird population and ecosystem monitoring. We briefly outline these research topics and describe their potential contribution to seabird monitoring.

SPECIFIC MONITORING PRIORITIES

- 1) For estimating breeding success, map colony boundaries whenever possible and use the estimators we provide in Section IV.
- 2) Consider if reproductive plots are representative of the colony as a whole. If in doubt, add some new, randomly selected plots in appropriate habitat within colony boundaries.
- 3) Archive all plot level data. Plot level data are much more useful than pooled data summaries when applying new statistical methods or assessing current sampling design.
- 4) Start assessing detectability (using Recommendation 1 in Section III) for each seabird that is monitored. Meeting this assumption is critical for attaining accurate counts of breeders. If complete detectability is not reasonable on a plot, it is certainly not reasonable for a total island-wide count (See Section III). For burrow nesters, we recommend a burrowscope for each island to determine burrow breeding activity.
- 5) Investigate the utility of using mark-recapture ratio estimator to adjust Mean incubation counts for renesting and double-clutching. Good candidates for pilot studies include Red-Footed Boobies and Red-Tailed Tropicbirds, because banded populations already exist. Dr. Ian Jones (Department of Biology, Memorial University, St. John's, Newfoundland A1B 3X9 CANADA; phone: (709) 737-7666; e-mail: ijones@mun.ca) is currently conducting banding studies with

Red-Tailed Tropicbirds, but does not monitor breeding success. We suggest that the USFWS consider collaborating with him to test the mark-recapture ratio estimator technique on Tern Island. Existing Red-tailed Tropicbird data from Midway Atoll may also be useful for investigating the mark-recapture approach.

GENERAL MONITORING RECOMMENDATIONS

- 1) If there is no plan to use banding information, we suggest not banding to avoid disturbance to birds and save resources. If managers plan to estimate apparent survival, apply the mark-recapture ratio estimator to correct MIC methods for asynchrony, or calibrate telomere methods, then we recommend banding individuals on reproductive plots.
- 2) When only one or two counts are possible each year, we recommend not using MIC methods. For example, because there is only one count per year on Laysan Island observers likely miss the peak breeding periods in many years, thus the counts are highly variable. As examples, see the accounts for Great Frigatebird and Red-footed Booby in Section II. To reduce the variability, the populations need to be surveyed more frequently.
- 3) Consider plot based studies, instead of island-wide counts. Because the entire island can be sampled, does not mean that it has to be. The simple random sample estimator provided in Recommendation 4, Section III is appropriate if detection is complete.
- 4) We suggest dedicating one person to assessment and application of monitoring protocols and management of monitoring data. Protocols will often be situation specific and will require some “trial and error” testing and some “tweaking” to be applied to different refuges.

RESEARCH PRIORITIES FOR SEABIRD MONITORING

Why research is important for seabird monitoring

In the introduction of this document, we outlined two monitoring goals. These goals were to (1) establish monitoring that can detect and understand changes in seabird populations, and (2) integrate seabird monitoring into an overall assessment of ecosystem health. In this document, we have used available data to evaluate the ability of population monitoring to detect long-term trends in population size and changes in reproductive success. For many species, we have demonstrated that there is a reasonable expectation of meeting these goals. However, for

other species, this may be more difficult. In some cases our ability to meet these goals is limited because the breeding colonies are inaccessible, and cannot be monitored on a regular basis. In other cases, we simply do not know enough about the breeding biology of particular species to predict how successful the application of monitoring methods will be. Finally, for some questions, especially those that use seabirds as indicators of ecosystem conditions, we may lack basic tools for linking seabirds to the ecosystem characteristics of interest. Here, we provide a brief summary of several areas of research that may have the potential to make significant contributions to seabird population monitoring and the use of seabirds as indicators of ecosystem conditions.

We still need more basic breeding biology information

One of the recurrent themes in this document is that we still lack some basic information on the breeding biology of many seabird species. This information, such as the probability of re-nesting after a breeding attempt fails, is an important component to interpreting counts of breeding seabirds (see Section I). We encourage managers to carefully consider the assumptions of their monitoring programs and whether there is sufficient information on the breeding biology of target species to justify these assumptions. If not, additional research is needed. One possibility may be to design these research questions into larger monitoring programs. Alternatively, there may be existing data that can be used to answer these questions. This document provides an example of how existing data can inform the design of monitoring programs.

Telomeres to estimate age structure

The age structure of a population is closely linked to patterns of survival. As noted in Section I, adult survival provides a large amount of information about the demographics of a population. Unfortunately, estimating age structure through banding of individual is labor intensive and may be a significant disturbance to individual birds. Telomere aging may provide an alternative way to collect information about age structure.

Telomeres are short, repeated DNA segments that cap the ends of eukaryotic chromosomes. They serve to stabilize chromosomes during replication and aid in chromosomal segregation. To measure telomere length, researchers collect red blood cells from birds and then use enzymes to cleave DNA into Telomere Restriction Fragments (TRFs). The relationship

between TRF and age is known as the “telomere clock” and is determined by regressing telomere length versus age for known aged (i.e., marked) individuals (e.g., Hausmann and Vleck 2002, Hausman et al. 2003, Hall et al. 2004).

If telomeres shorten predictably with age, then telomeres can be measured in tissue samples collected from birds, and un-marked individuals can be classified into age classes. If the age distribution of a population can be determined with enough precision, an island could (in theory) be visited only periodically and the growth rate of the population could be determined. At this time, however, there is not enough information to verify that telomere measurements can be used as an effective monitoring tool. In fact, in a recent paper on using telomeres to age Great Frigatebirds at French Frigate Shoals, Juola et al. (2006) concluded that “estimations of age of individuals and of estimated age structure of breeding birds in this population are not particularly reliable.”

Telomere aging is still experimental, and not a proven monitoring tool at this time, however, we recognize the development of such a tool would be a great advance in seabird monitoring. The Northwest Hawaiian Islands, with many populations of banded birds of known age, provide an excellent opportunity to further research the relationship between telomere length and age. A contact for more information on telomere length is Dr. Carol Vleck (Department of Zoology and Genetics, Iowa State University, Ames, IA 50014; (515) 294-8646; email: cvleck@iastate.edu).

Corticosterone as an indicator of food availability

Reproductive success is widely used as a metric of food availability and other conditions experienced by birds: when food is widely available, reproductive success is high. This relationship is the core of the concept that seabirds can act as indicators of ecosystem conditions. Measuring chick growth rates can also provide information about food availability. Unfortunately, measuring reproductive success or growth rates is labor intensive and requires an extensive time commitment. Corticosterone levels may correspond with food availability and be much easier to collect than reproductive or growth rate data.

Corticosterone (CORT) is a hormone that is secreted by the adrenal gland of birds when they are stressed. The secretion of CORT is important for nest-bound birds, because it leads to

elevated levels of food begging and more aggressive behavior (Kitaysky et al. 2001, 2003). Hence, CORT appears to act as a mechanism by which birds alter their behavior in response to food shortages. Food shortage leads to elevated CORT and, in chicks, elevated CORT leads to increased begging and more aggression (in adults the response is different). The link between food availability and CORT levels has been verified in both experimental (Kitaysky et al. 1999a, 2003) and observational studies (Kitaysky et al. 1999b). Many seabirds respond to food stress (i.e., food shortages) by secreting increased levels of CORT (Nunez-de la Mora et al. 1996; Kitaysky et al. 1999a, 1999b; Kitaysky et al. 2001).

CORT is measured from blood samples of adult or nestling birds. Biologists capture a bird, puncture the alar vein, and collect a blood sample. Birds are held for about an hour, and blood samples are collected at 3, 10, 30, and 50 minutes. This time-series of measurements is referred to as a “stress series”. CORT levels build as the birds are held captive. The stress series is used to determine maximum CORT levels. Based upon data from Alaska, it appears that sampling 10 birds is sufficient during each sample period. More details on sampling design are given in Wingfield et al. (1982). Maximum CORT levels reflect food stress over a period of weeks (but this period varies among species). As such, the method may provide a precise tool for monitoring food availability. For instance, CORT measurements may identify periods of food deprivations even when reproductive success is unaffected. As such, they may be useful for monitoring the effects of periodic or low-level reductions in food availability.

Because CORT may provide an efficient measurement of food availability for nesting seabirds, we recommend supporting research on corticosterone monitoring. However, implementing CORT research into a monitoring program will require techniques be tested and calibrated. As such, we recommend that one or two species be selected for pilot studies on how food availability relates to CORT. Dr. Sasha Kitaysky is a contact for more information or advice (ffask@uaf.edu; 907-474-5179).

Methods for monitoring diet composition

In addition to information on how much food is available (measured by reproductive success or corticosterone levels), managers may be interested in what types of food are available. Traditionally, this has been conducted with regurgitation samples. While such methods are commonly used, they are likely biased. First, regurgitations or stomach analyses examine only

one meal. Second, due to digestion, small prey items or small prey species may be under-represented. Third, because different prey species have different digestive rates, species with hard parts may be over-represented. Last, adults may consume different prey items than what they bring to the nest to feed their chicks. If regurgitations are conducted at the colony, a biased view of diet is presented (i.e., you only see what they bring their chicks). Two alternatives to regurgitation samples are available (1) fatty acid signature analysis, and (2) stable isotope analysis.

Quantitative fatty acid signature analysis. This technique uses the composition of fatty acids to assess seabird diet composition (Iverson et al. 2004). Fatty acids are the main component of most lipids (i.e., the fat in prey species). These fatty acids are released during digestion and do not break down like proteins. In short, fatty acids from prey are digested, taken into circulation, and deposited into tissues in an intact and recognizable form. If the fatty acid signature in prey species is known, one can assay the fatty acids in seabirds to determine what individuals are feeding upon and what proportion of their diet is composed of different prey species.

Biologists collect food samples from birds via regurgitation. These samples serve as the base “signature” for different prey species. Fat biopsies must also be collected from living birds. The bird is captured and a small incision (approx. 1 cm in length) is made near the base of the tail. A small amount of fat is removed from this incision and this fat is either frozen or stored in a preservative solution. The incision is then closed with a skin adhesive. The technique is being widely applied in the Alaska Maritime National Wildlife Refuge by Sara Iverson and Alan Springer, and complications for individual birds are rarely reported (Iverson and Springer 2002).

Stable isotope analysis. Like fatty acids, stable isotopes may provide another means of collecting information about foraging patterns and diet composition. The ratios of isotopes of carbon ($^{13}\text{C}/^{12}\text{C}$) and nitrogen ($^{15}\text{N}/^{14}\text{N}$) have been shown to vary in seabirds with both trophic position and foraging location (nearshore vs. pelagic) (Sydeman et al. 1997). The application of stable isotope analysis to seabird biology has been relatively well studied (Stapp et al. 1999, Hobson et al. 1994, Forero and Hobson 2003). Thus, we feel that research questions concerned with trophic position and/or foraging locations of seabirds may be effectively addressed with this technique.

If seabird managers are interested in collecting information about diet composition, we encourage research into the application of quantitative fatty acid signature analysis and stable isotopes. Although these methods may be more expensive than collecting regurgitations alone, we suggest that they are an important component to quantifying diet. Alan Springer (Institute of Marine Science, 262 Arctic Health, University of Alaska Fairbanks, Fairbanks, AK 99775-1080; Phone: (907) 474-6213; email: ams@ims.uaf.edu) and Sara Iverson (Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1; email: sara.iverson@dal.ca) are contacts for research on fatty acid signature analysis.

LITERATURE CITED

- Bart, J., and S. Earnst. 2002. Double sampling to estimate density and population trends in birds. *Auk* 119:36-45.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling*. Oxford University Press, New York, New York.
- Burgman, M. A., S. Ferson, and H. R. Akcakaya. 1993. *Risk assessment in conservation biology*. Chapman and Hall, London, UK.
- Burnham, K. P., and D. R. Anderson. 1998. *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New York, New York.
- Burnham, K. P., D. R. Anderson, G. C. White, C. Brownie, and K. H. Pollock. 1987. *Design and analysis methods for fish survival experiments based on release-recapture*. American Fisheries Society Monograph No. 5. American Fisheries Society, Bethesda, Maryland.
- Byrd, G. V., D. I. Moriarty, and B. G. Brady. 1983. Breeding biology of Wedge-tailed Shearwaters at Kilauea Point, Hawai'i. *Condor* 85:292-296.
- Chapman, D. G. 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. *University of California Publication of Statistics* 1:131-160.
- Clarke, E. B., L. B. Spear, M. L. McCracken, F. F. C. Marques, D. L. Borchers, S. T. Buckland, and D. G. Ainley. 2003. Validating the use of generalized additive models and at-sea surveys to estimate size and temporal trends of seabird populations. *Journal of Applied Ecology* 40:278-292.
- Cochran, W. G. 1977. *Sampling techniques*, 3rd edition. John Wiley & Sons, Inc., New York, New York.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Academic Press, New York, New York.
- Dearborn, D. C., and A. D. Anders. 1996. Reproductive success of black noddies, red-footed boobies, and red-tailed tropicbirds on Tern Island, French Frigate Shoals in 1995. U.S. Fish and Wildlife Service, Pacific – Remote Islands NWR, Pacific Region, Honolulu, Hawai'i.
- Dearborn, D. C., A. D. Anders, and E. N. Flint. 2001. Trends in reproductive success of Hawaiian seabirds: Is guild membership a good criterion for choosing indicator species? *Biological Conservation* 101:97-103.
- Doherty, Jr., P.F., E. A. Schreiber, J. D. Nichols, J. E. Hines, W. A. Link, G. A. Schenk, and R. W. Schreiber. 2004. Testing life history predictions in a long-lived seabird: a population matrix approach with improved parameter estimation. *Oikos* 105:606-618.

- Dyer, P. K., and G. J. E. Hill. 1991. A solution to the problem of determining the occupancy status of Wedge-tailed Shearwater *Puffinus pacificus* burrows. *Emu* 91:20-25.
- Fefer, S. C. Harrison, M. B. Naughton, and R. J. Shallenberger. 1984. Synopsis of results of recent seabird research conducted in the Northwestern Hawaiian Islands. Pp. 9-76 in *Proceedings of the second symposium on resource investigations on Northwestern Hawaiian Islands, Volume 1, Honolulu, Hawai'i.*
- Forero, M. G., and K. A. Hobson. 2003. Using stable isotopes of nitrogen and carbon to study seabird ecology: applications in the Mediterranean seabird community. *Scientia Marina* 67:23-32.
- Frederick, P. C., J. A. Heath, R. Bennetts, and H. Hafner. 2006. Estimating nests not present at the time of breeding surveys: an important consideration in assessing nesting populations. *Journal of Field Ornithology* 77:212-219.
- Gauger, V. H. 1999. Black Noddy. In A. Poole and F. Gill, editors, *The Birds of North America*, No. 412. The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington DC.
- Gerrodette, T. 1987. A power analysis for detecting trends. *Ecology* 68:1364-1372.
- Gerrodette, T. 1991. Models for power of detecting trends: a reply to Link and Hatfield. *Ecology* 72:1889-1892.
- Gerrodette, T. 1993. Program TRENDS: user's guide. Available on the web at: swfsc.nmfs.noaa.gov/PRD/software/trends_guide.pdf.
- Gould, W. R., and J. D. Nichols. 1998. Estimation of temporal variability of survival in animal populations. *Ecology* 79:2531-2538.
- Hall, M. E., L. Nasir, F. Daunt, E. A. Gault, J. P. Croxall, S. Wanless, and P. Monaghan. 2004. Telomere loss in relation to age and early environment in long-lived birds. *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:1571-1576.
- Hatch, S. A. 2003. Statistical power for detecting trends with applications to seabird monitoring. *Biological Conservation* 111:317-329.
- Harding, A. M. A., J. F. Piatt, G. V. Byrd, S. A. Hatch, N. B. Konyukhov, E. U. Golubova, and J. C. Williams. 2005. Variability in colony attendance of crevice-nesting horned puffins: Implications for population monitoring. *Journal of Wildlife Management* 69:1279-1296.
- Harrison, C. S., T. S. Hida, and M. P. Seki. 1983. Hawaiian seabird feeding ecology. *Wildlife Monographs* 85:1-71.
- Hausmann, M. F., and C. M. Vleck. 2002. Telomere length provides a new technique for aging animals. *Oecologia* 130:325-328.

- Hausmann, M. F., D. W. Winkler, K. M. O'Reilly, C. E. Huntington, I. C. T. Nisbet, and C. M. Vleck. 2003. Telomeres shorten more slowly in long-lived birds and mammals than in short-lived ones. *Proceedings of the Royal Society of London, Series B* 270:1387-1392.
- Hobson, K. A., J. F. Piatt, J. Pitocchelli. 1994. Using stable isotopes to determine seabird trophic relationships. *Journal of Animal Ecology* 63:786-798.
- Iverson, S. J., and A. M. Springer. 2002. Estimating seabird diets using fatty acids: protocol development and testing of ReFER hypotheses. Final Report to the National Pacific Marine Research Program. University of Alaska, Fairbanks, Alaska.
- Iverson, S. J., C. Field, W. D. Bowen, and W. Blanchard. 2004. Quantitative fatty acid signature analysis: a new method of estimating predator diets. *Ecological Monographs* 74:211-235.
- Juola, F. A., M. F. Hausmann, D. C. Dearborn, and C. M. Vleck. 2006. Telomere shortening in a long-lived marine bird: cross-sectional analysis and test of an aging tool. *Auk* 123:775-783.
- Kish, L. 1965. Survey sampling. John Wiley & Sons, Inc., New York, New York.
- Kitaysky, A. S., and E. G. Golubova. 2000. Climate change causes contrasting trends in reproductive performance of planktivorous and piscivorous alcids. *Journal of Animal Ecology* 69:248-262.
- Kitaysky, A. S., E. V. Kitaiskaia, J. F. Piatt, and J. C. Wingfield. 2003. Benefits and costs of increased levels of corticosterone in seabird chicks. *Hormones and Behavior* 43:140-149.
- Kitaysky, A. S., J. C. Wingfield, and J. F. Piatt. 1999b. Dynamics of food availability, body condition and physiological stress response in breeding Black-legged Kittiwakes. *Functional Ecology* 13:577-584.
- Kitaysky, A. S., J. F. Piatt, J. C. Wingfield, and M. Ramano. 1999a. The adrenocortical stress-response of Black-legged Kittiwake chicks in relation to dietary restrictions. *Journal of Comparative Physiology B* 169:303-310.
- Kitaysky, A.S., E. V. Kitaiskaia, J. C. Wingfield, and J. F. Piatt. 2001. Dietary restriction causes chronic elevation of corticosterone and enhances stress response in red-legged kittiwake chicks. *Journal of Comparative Physiology B*. 171:701-709.
- Lancia, R. A., J. D. Nichols, and K. H. Pollock. 1996. Estimating the number of animals in wildlife populations. Pages 215-253 *in* T. A. Bookhout, editor, *Research and management techniques for wildlife and habitats*, 5th edition. The Wildlife Society, Bethesda, Maryland.
- Link, W. A., and J. D. Nichols. 1994. On the importance of sampling variance to investigations of temporal variation in animal population size. *Oikos* 69:539-544.

- Link, W. A., and J. R. Sauer. 1997. Estimation of population trajectories from count data. *Biometrics* 53:488-497.
- Link, W. A., and J. S. Hatfield. 1990. Power calculations and model selection for trend analysis: a comment. *Ecology* 71:1217-1220.
- Link, W. A., and J. R. Sauer. 1998. Estimating population change from count data: Application to the North American Breeding Bird Survey. *Ecological Applications* 8:258-268.
- Little, R. C., W. W. Stroup, and R. J. Freund. 2002. SAS® for linear models, 4th edition. SAS Institute, Cary, North Carolina.
- Lenth, R. V. 2001. Some practical guidelines for effective sample size determination. *The American Statistician* 55:187-193.
- Massey, B. A., and J. L. Atwood. 1981. Second-wave nesting of the California Least Tern: age composition and reproductive success. *Auk* 98:596-605.
- Megyesi, J. L., and D. L. O'Daniel. 1997. Bulwer's Petrel. *In* A. Poole and F. Gill, editors, *The Birds of North America*, No. 281. The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington DC.
- Mendenhall, W., L. Ott, and R. L. Scheaffer. 1971. *Elementary survey sampling*. Wadsworth Publishing Company, Inc. Belmont, California.
- Montevicchi, W. A. 1993. Birds as indicators of change in marine prey stocks. Pages 217-266 *in* Furness, R. W., and J. J. D. Greenwood, editors, *Birds as Monitors of Environmental Change*. Chapman and Hall, London.
- Nichols, J. D. 1991. Extensive monitoring programs viewed as long-term population studies: The case of North American waterfowl. *Ibis Supplement* 1:89-98.
- Nunez-de la Mora, A., H. Drummond, and J. C. Wingfield. 1996. Hormonal correlates of dominance and starvation-induced aggression in chicks of the Blue-footed Booby. *Ethology* 102:748-761.
- Polovina, J. J., G. T. Mitchum, N. E. Graham, M. P. Craig, E. E. Demartini, and E. N. Flint. 1994. Physical and biological consequences of a climate event in the central North Pacific. *Fisheries Oceanography* 3:15-21.
- Raj, D. 1968. *Sampling theory*. McGraw-Hill, New York, New York.
- Rauzon, M. J., C. S. Harrison, and R. B. Clapp. 1984. Breeding Biology of the Blue-Gray Noddy. *Journal of Field Ornithology* 55:309-321.
- Russell, R. W. 1999. Comparative demography and life history tactics of seabirds: Implications for conservation and marine monitoring. *American Fisheries Society Symposium* 23:51-76.

- Saether, B. E., and O. Baake. 2000. Avian life history variation and contribution of demographic traits to the population growth rate. *Ecology* 81:642-653.
- Scheaffer, R. L., W. Mendenhall, and R. L. Ott. 1996. Elementary survey sampling, 5th edition. Wadsworth Publishing Company, Belmont, California.
- Schreiber, R. W. 1980. The Brown Pelican: An endangered species? *Bioscience* 11:742-747.
- Seto, N. W. H. 2001. Christmas Shearwater. *In* A. Poole and F. Gill, editors, *The Birds of North America*, No. 561. The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington DC.
- Seto, N. W. H., and D. O'Daniel. 1999. Bonin Petrel. *In* A. Poole and F. Gill, editors, *The Birds of North America*, No. 385. The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington DC.
- Small, C. 1999. Bonin Petrel burrow density mapping 1999 - Sand Island. Unpublished Report. Midway Atoll NWR Administrative Report, USFWS.
- Sokal, R. R. and F. J. Rohlf. 1995. *Biometry*, 3rd edition. W. H. Freeman and Company, New York, New York.
- Stapp, P., G. A. Polis, F. S. Piñero. 1999. Stable isotopes reveal strong marine and El Niño effects on island food webs. *Nature* 401:467-469.
- Steidl, R. J., J. P. Hayes, and E. Schaubert. 1997. Statistical power in wildlife research. *Journal of Wildlife Management* 61:270-279.
- Sydeman, W. J., K. A. Hobson, P. Pyle, and E. B. McLaren. 1997. Trophic relationships among seabirds in central California: combined stable isotope and conventional dietary approach. *Condor* 99:327-336.
- Taylor, L. L., and T. Gerrodette. 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conservation Biology* 7:489-500.
- Thomas, L., and K. Martin. 1996. The importance of analysis method for breeding bird survey population estimates. *Conservation Biology* 10:479-490.
- Thompson, S. K. 2002. *Sampling*, 2nd edition. John Wiley & Sons, Inc., New York, New York.
- Thompson, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. *Auk* 119:18-25
- Thompson, W. L., G. C. White, and C. Gowan. 1998. *Monitoring vertebrate populations*. Academic Press, Inc., San Diego, California.

- U.S. Fish and Wildlife Service. 1980. Monitoring manual – reproductive success study – general methods. U.S. Fish and Wildlife Service, Pacific – Remote Islands NWR, Pacific Region, Honolulu, Hawai`i.
- U.S. Fish and Wildlife Service. 2002. Birds of conservation concern 2002. Division of Migratory Bird Management, Arlington, Virginia.
- U.S. Fish and Wildlife Service. 2005. Regional seabird conservation plan, Pacific Region. U.S. Fish and Wildlife Service, Migratory Birds and Habitat Programs, Pacific Region, Portland, Oregon.
- Usher, M. B. 1991. Scientific requirements of a monitoring programme. Pages 15-32 *in* B. Goldsmith, editor. Monitoring for conservation and ecology. Chapman and Hall, London, UK.
- Vanderlip, C. 2004. Draft Northwestern Hawaiian Islands/Kure Atoll Field Report. Hawai`i Department of Land and Natural Resources, Division of Forestry and Wildlife, Honolulu, Hawai`i.
- Vanderlip, C., and A. Marie. 2003. Draft DLNR/DOFAW Kure Atoll Field Season: May - September 2003 Field Report. Hawai`i Department of Land and Natural Resources, Division of Forestry and Wildlife, Honolulu, Hawai`i.
- White, G. C., A. B. Franklin, and T. M. Shenk. 2002. Estimating parameters of PVA models from data on marked animals. In: Population viability analysis. S. R. Beissinger and D. R. McCullough, Eds. University of Chicago Press, Chicago, Illinois.
- Whittow, G. C. 1997. Wedge-tailed Shearwater. *In* A. Poole and F. Gill, editors, The Birds of North America, No. 305. The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington DC.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and Management of Animal Populations: Modeling, Estimation, and Decision Making. Academic Press, San Diego, California.
- Wingfield, J.C., J. P. Smith, and D. C. Farner. 1982. Endocrine responses of white-crowned sparrows to environmental stress. *Condor* 84:399–409.
- Woodward, P. W. 1972. The natural history of Kure Atoll, Northwestern Hawaiian Islands. *Atoll Research Bulletin* 164:1-318.
- Zielinski, W. J., and H. B. Stauffer. 1996. Monitoring marten populations in California: survey design and power analysis. *Ecological Applications* 6:1254-1267.

APPENDIX A. SAS CODE FOR MONTE CARLO POWER SIMULATIONS

We wrote a number of programs to handle specific situations, but only present the base code for one program. This program assumes the user has log transformed the data and that the variance is normally distributed on a log scale. In some instances, log transformation was not appropriate (e.g., Red-tailed Tropicbird data was not characterized by high variance when counts were large). This program also assumes a multiplicative trend (additive on a log scale). Most changes are easily implemented with a few alterations of code. Contact John Citta for more information.

```
/* ONE SAMPLE POWER ANALYSIS WITH CONSTANT VARIANCE ON LOG SCALE */
/*
/* This program generates univariate tests for prospective power using log transformed */
/* data assuming a linear model and normal error distribution. This model only applies */
/* for situations where there is one count per year and the count is assumed to be */
/* without error. */
/*
/* Inputs include mean, stdev, and trend - get data from standard GLM on log trans- */
/* formed data. For each length of time, the array "YEAR" and the do-loops must be */
/* altered to the correct length. */
/*
/* There are separate data and array steps, because the regression should start at year */
/* 0, but arrays must be defined from 1 to n. Hence, a column of years, starting at 0 */
/* must be added to the array of trend values. */
/*
/* The baseline trend is computed in the array "N" and this array is modified to */
/* simulate stochastic variation in the array "YEAR". */
/*
/* Output is into the table "TEMP" in the SAS directory "WORK" - P-values for the */
/* intercept and slopes are added to this table. Sort on the slope (i.e., variable */
/* "YEAR") and calculate percentage of tests that have significant slopes. */

/* Written by John Citta, October 2005.*/

dm 'clear output';
dm 'clear log';
LIBNAME SASDATA 'C:\SASDATA\';
PROC PRINTTO LOG='C:\SASDATA\LOGFILE.TMP';
DATA YEAR;
INPUT YEAR; /*For comparison with TRENDS, the correct number of years is the number of years */
/*in the DATA step, not the ARRAYS below!! */
CARDS; /*the "year" column - this must equal duration of monitoring in years starting at 0*/
0
1
2
3
4
5
```

```

6
7
8
9
10
;
/*PROC PRINT DATA=YEAR; */      /*proc for data checking - normally turned off*/
/*RUN;*/

%MACRO REG;                        /*The simulations and tests are embedded in this macro */
%DO REP=1 %TO 2500;                /*this is the number of simulations to run */
Data DAT (DROP=TREND DROP=START DROP=SDEV DROP=R DROP=I DROP=X1);
    TREND=-0.06931;                 /*-0.06931 IS THE ADDITIVE TREND ON A LOG SCALE */
                                    /*THIS IS THE CORRECT VALUE FOR A 50% DECLINE OVER 10 */
                                    /*YEARS ON A NON-LOG SCALE – on a non-log scale, the trend is
                                    /*-0.06697*/

    START=6.9078;                  /*STARTING VALUE ON LOG SCALE */
    SDEV=0.234651;                 /*SDEV IS STANDARD DEVIATION FROM DETRENDED LOG */
                                    /* TRANSFORMED DATA*/
                                    /*THIS IS EQUAL TO "MSE" IN PROC GLM*/

DO R=1 TO 1;
ARRAY N [11]_TEMPORARY_; /*THIS MUST BE CHANGED FOR EACH DURATION OF*/
                            /*MONITORING */

    N(1)=START;
    DO I=2 TO 11;                /*THIS MUST BE CHANGED FOR EACH DURATION OF
                                /*MONITORING */
        N(I)=(N(I-1)+TREND); /*ON A LOG SCALE THE MULTIPLICATIVE TRENDS ARE
                                /*ADDITIVE!!! */
    END;                          /* For a non-transformed scale the code is: */
                                    /* "N(I)=((N(I-1)*TREND)+N(I-1))" */

ARRAY YR [11];                  /*THIS MUST BE CHANGED FOR EACH DURATION OF
                                /*MONITORING*/

    YR(1)=START;                 /*This is the starting value*/
    DO I=2 TO 11;                /*THIS MUST BE CHANGED FOR EACH DURATION OF
                                /*MONITORING*/

        X1=N(I);
        X1=X1+(SDEV*NORMAL(0)); /*This is the random number generator - TO CHECK
                                /* THAT THE CODE IS PRODUCING THE CORRECT
                                /* TREND, TURN OFF THIS LINE OF CODE AND
                                /* EXAMINE THE OUTPUT ARRAY*/

        IF X1>1 THEN X1=X1; /*This creates a boundary - the log of 0 is undefined*/
        ELSE X1=1;

        YR(I)=X1;
    END;
    OUTPUT;
END;

PROC TRANSPOSE OUT=D2;
RUN;

DATA ALL;
MERGE YEAR D2;
RUN;

```

```

/*PROC PRINT DATA=ALL;    */           /*proc for data checking - normally turned off*/
/*RUN;*/

PROC PRINTTO PRINT=TEMP;           /*This prints the file to work.junk -
the file does not erase between sims*/
RUN;                               /*Be certain to delete file between sims*/

PROC GLM DATA=ALL;               /*This is the statistical test - currently
written to output values to table "temp"*/

CLASS NAME;
MODEL COL1=YEAR / SOLUTION;
ODS EXCLUDE MODELANOVA;
ODS OUTPUT PARAMETERESTIMATES=PARMS;
RUN;

PROC APPEND BASE=TEMP DATA=PARMS; RUN;
RUN;

%END;
%MEND REG;

%REG;

```


APPENDIX B. COMPARISON OF PROGRAM TRENDS AND MONTE CARLO SIMULATIONS

TRENDS (Gerrodette 1987, 1991) is a widely available software package for analyzing trends in count data. A major limitation of TRENDS is that only one form of variation can be incorporated into calculations of power (Hatch 2003). The user can account for annual variation or observation error, but not both. Furthermore, TRENDS only allows the user to account for three functional relationships between variance and population size. We wrote a Monte Carlo program in SAS that allows the user to specify the number of within season counts and the level of variation in those counts. This code can be easily altered to allow for any relationship between variance and population size that is observed. To test the base code, we parameterized the simulations to allow for only one count per season and then compared this directly to program TRENDS. As an example, we use Black Noddy data from Tern Island between 1980 and 2003. Input for both programs is on a log scale. Simulations assume a standard deviation of 0.5539, an exponential trend of -0.0694, and a one-sided significance test with $\alpha = 0.05$. In both programs, power is derived with t-tests. In our Monte Carlo simulation, power is equal to the proportion of significant tests in 2500 simulations. The resulting estimates of power are virtually identical (Table B.1; Fig. B.1) when there is no serial autocorrelation in counts.

Table B.1. A comparison of statistical power between program TRENDS and our Monte Carlo method.

Duration of sampling in years	Power (1- β)	
	TRENDS	Monte Carlo
5	0.10	0.10
6	0.13	0.13
7	0.16	0.17
8	0.20	0.21
9	0.25	0.27
10	0.32	0.34
11	0.39	0.42
12	0.47	0.50
13	0.55	0.59
14	0.64	0.67
15	0.72	0.74
16	0.79	0.81
17	0.85	0.87
18	0.90	0.92
19	0.94	0.94
20	0.97	0.97

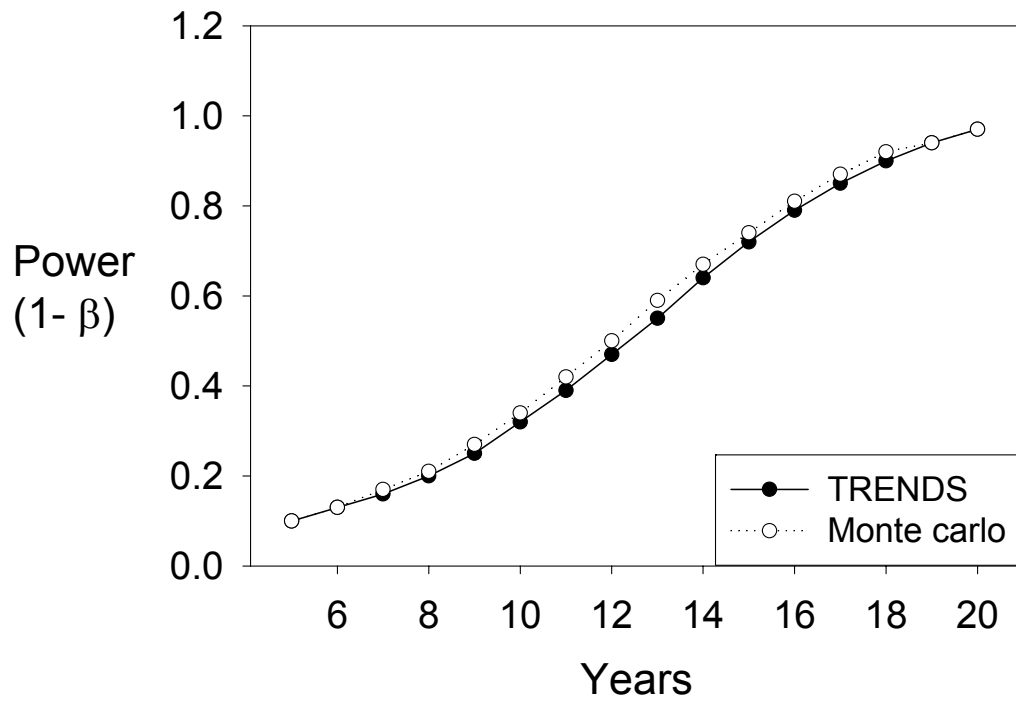


Figure B.1. A comparison of statistical power between program TRENDS and our Monte Carlo method.

APPENDIX C. OUTLINE OF MIC ALTERNATIVES

This is a list of alternatives for MIC methods that we considered. For each alternative, we assessed what methods would be necessary, the advantages to using the alternative, the disadvantages to using the alternative, and our assessment of the alternative. We have included general information about the type of data that can be generated by each method, the logistical constraints and resources that would be required, and the possibility for disturbance to the nesting colonies. We recognize, however, that the relative importance of these factors will vary among islands. Thus, the choice of the most appropriate metric may vary depending on the monitoring objectives, resource availability, logistical constraints, and disturbance risks for different islands or monitoring programs. Here, we present information on the strengths, weaknesses, and assumptions of some alternative metrics to facilitate the decision making process by FWS staff that are familiar with the conditions at the breeding colonies.

1. Use alternative metrics that are pre-existing:

- a. Method: A number of metrics are collected in addition to MICs. These data include the total number of nests, the total number of nests with eggs, and the total number of nests with chicks. There are a number of reasons to suspect these are better metrics than MICs. If few Black Noddies renest or relay in a given year, then the total count of nests with eggs may be a better correlate of breeding population size and may be less variable. Likewise, if few Black Noddies double-brood, but all will renest until they produce one successful clutch, then the total count of nests with chicks may be a better correlate of population size and may be less variable.
- b. Advantages:
 - i. If alternative metrics are more correlated to true population size, they will likely exhibit less variation and statistical power will be higher.
 - ii. These data are already collected, so there will be no new methods.
 - iii. Historical data can be used to reassess power.
- c. Disadvantages:
 - i. We must still assume a consistent relationship between the alternative metric and the number of breeding birds.
 - ii. There are still no statistically valid measures of precision within years.
- d. Assessment: For select species (e.g., BLNO) the use of alternative metrics was assessed by comparing the variance of MIC with the variance of total nests, total nests with eggs, and total nests with chicks. We then calculated the statistical power associated with levels of variation (see species accounts). Because there is still no statistically valid measure of precision associated with using these metrics, because we still must assume a consistent relationship between the metric and the true number of breeding birds, and because the use of alternative metrics did not greatly improve statistical power for any species, we conclude that this will often not be the best alternative.

2. Conduct multiple counts per incubation period:

- a. Method: Breeding pairs could be counted more than once per incubation period. This effectively increases the sample size. Instead of having a sample of 1 count per calendar year, there would be two or more.
- b. Advantages:
 - i. Having more than one count in a season is expected to increase statistical power.
 - ii. The same methods can be used.
- b. Disadvantages:
 - i. Counts are expected to be highly autocorrelated. Although our analyses indicated that there was no autocorrelation between years, we think it is virtually impossible that repeat counts, within an incubation period, will not be autocorrelated. If a census is plausible, then counts within incubation periods will be virtually 100% correlated. In order to correctly estimate the increase in power due to multiple counts within incubation periods, analyses will have to account for an autocorrelation structure where there is high autocorrelation within years, but little between years. While such analyses are not computationally difficult, the ability of our analyses to detect such complicated forms of autocorrelation is likely low.
 - ii. Observers must still assume a consistent relationship between the additional counts and the true number of breeding birds.
 - iii. This method still does not account for the probability of detection.
 - iv. This method will cause more disturbance over having a single count.
 - v. This method will require more time investment than having a single count.
- c. Assessment: Because repetitive counts within incubation periods are expected to introduce complex forms of autocorrelation into the data set, because the ability of the methods we used to detect autocorrelation structure likely cannot detect such complicated forms of autocorrelation, and because this method is not likely to produce estimates of breeding population size that are closer to the true breeding population size (i.e., this method inherently has all the problems that standard, single-count MIC methods have), we conclude it will often not be the best alternative.

3. Account for detectability:

- a. Method: Detectability, or the probability of detection, is the likelihood that a subject is detected, given it is present. For counts of nests, the subject is the nest. It is unlikely that all nests are detected in counts. If detectability were constant through time, then it could be ignored. However, detectability likely changes with observer experience, the density of nests, and the distance an observer is from a nest. If detectability is not accounted for, stable populations may appear variable. By accounting for detectability, variance may decline and statistical power may increase.
- b. Advantages:
 - i. Variance in counts will likely decrease and power will increase.
 - ii. This method will yield statistically defensible measures of nest abundance.
- c. Disadvantages:
 - i. This method will require more time investment.
 - ii. This method would be intrusive, a sample of nests would have to be marked each year (unlike marked individuals that would remain marked).
 - iii. We would still have no insight regarding the proportion of birds that are double counted in the MIC and cannot assume a consistent relationship between the count and the number of breeders.
- d. Assessment: We recommend that detectability be assessed for all species. However, detectability is likely a minor issue when compared to the fundamental problems associated with MIC methods. In other words, asynchrony likely causes much more variation in count

data than variation in detectability. However, we still recommend that detectability be addressed. See Section III for a complete discussion of detectability.

4. Mark-recapture with nests:

- a. Method: This is a closely related technique to that described above (#2). The total number of nests can be directly quantified by estimating the probability of encounter and model the abundance of nests with a mark-recapture model. Nests could be marked and then resighted within each sample period. This would directly estimate the probability of encounter and if the method was repeated over time, we could directly model trends within these data.

Closed mark-recapture models (e.g., Lincoln-Peterson) would be used to estimate the total number of nests. Observers would mark nests and estimate detectability (p_i). Where N_i is total number of nests and O_i is the number of nests observed:

$$\hat{N}_i = \frac{O_i}{p_i}.$$

- b. Advantages:
 - i. Statistically defensible measures of nest abundance and well developed methods to estimate trends.
- c. Disadvantages:
 - i. This method will require more time investment.
 - ii. This method would be intrusive, a sample of nests would have to be marked each year (unlike marked individuals that would remain marked).
 - iii. We would still have no insight regarding the proportion of birds that are double counted in the MIC and cannot assume a consistent relationship between the count and the number of breeders.
- d. Assessment: As described in Alternative 3 above, accounting for detectability is important. When monitoring species that do not nest more than once in a season, nest detectability is probably the next most important issue. This method is a relatively easy way to address detectability. However, it does not circumvent the main issue with mean incubation counts when pairs nest more than once in a season or renest after failure – that we do not know how the minimum count relates to the true number of breeders. Without marked birds (as described in Alternative 7), we still cannot derive an accurate count of the number of breeders.

5. Distance sampling with nests:

- a. Method: Distance sampling (Buckland et al. 1993) is another method where a partial count can be used to estimate the total number of nests. Within distance sampling, either points or lines are the sampling unit. The distance to individual nests would be recorded and this information is used within program DISTANCE to generate a “sightability function”, assuming that individuals on the transect line have complete detectability. The software uses this function to estimate the density of individuals along points or line transects. If the total area a species occupies is known, then this can be converted into estimates of abundance.
- b. Advantages:
 - i. This method yields statistically defensible measures of nest abundance
 - ii. The theory and software are well developed
- c. Disadvantages:
 - i. Still have no insight on the proportion of birds that are double counted in the MIC. Cannot assume a consistent relationship between the count and the number of breeders.

- ii. Distance sampling is usually used when the entire area cannot be sampled. This is not the case on most islands.
- iii. If observers can get close enough to nests to have a complete count, then Distance Sampling is not necessary.
- iv. Monthly estimates of the number of nests based upon Distance Sampling still cannot be related to the total number of breeders and therefore still has the fundamental limitation of mean incubation counts.
- v. This method will be more preferred for cryptic species or where the entire area cannot be sampled.
- d. Assessment: In general, this method will not be preferred. In instances where the power of MIC methods to detect trends is low, Distance Sampling is not likely to decrease variance. Rather, distance sampling will be useful in situations where the entire island cannot be sampled. However, the method will have the same limitations as MICs.

6. Mark-recapture of individual birds (not nests):

- a. Method: Banding birds with combinations of bands that allow identifying individuals via resighting.
- b. Advantages:
 - i. By marking and resighting individual birds, we can determine the total number of breeding birds within a season (i.e., this circumvents all problems associated with mean incubation counts).
 - ii. We can also gain ancillary information that will be useful for understanding the basic ecology of these species and allow the parameterization of population models. For example, ancillary information would include survival rates, breeding propensity, renesting propensity, etc.
 - iii. Mark recapture data would yield estimates of total population size and/or population rate of change (i.e., not just number of breeding birds).
 - iv. All issues of detection probability would be accounted for.
 - v. Mark-recapture methods have a well developed theory and methods.
- c. Disadvantages:
 - i. Mark-recapture studies will likely cause more disturbance.
 - ii. Mark-recapture studies will be more expensive.
 - iii. The refuge would likely have to find additional expertise to analyze data
- d. Assessment: Large-scale mark-recapture studies provide the most information, but they are likely not possible or desired for all species. The analytical tools (e.g., specialized software) for these projects are developing rapidly and require substantial training to be used effectively. An introduction to some of the analytical tools, data collection protocols, and logistical challenges for such studies is available in a complementary report on demographic monitoring of Laysan and Black-footed Albatross in the Hawaiian Islands (M. B. Naughton, pers. comm.). When such an approach is infeasible, we suggest integrating limited mark-recapture methods with current counting methods is a feasible compromise (See Alternative 7, below).

7. Mark-recapture on study plots only:

- a. Method: The general idea is to use information from currently monitored study plots to determine how the number of total breeders relates to mean incubation counts. In effect, we compute a conversion factor on the reproductive plots and apply this to the entire population. This is accomplished with a *ratio estimator*.

The theory comes from Cochran (1977) and Thompson (1992). Assume the island consists of N possible plots, of which n are sampled. Also assume that the total population count is M_{total} ; this is the total number of nests (nests with eggs) counted on the island. The

plot level counts are m_i for plots 1- n (i.e., plot 1 is m_1 , plot 2 is m_2 , etc.). On each plot, mark all birds. The number of unique pairs on each plot equals y_i for plots $1-i$. The ratio of interest is:

$$r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n M_i}$$

The number of breeders is estimated as:

$$\# \text{ pairs} = rM_{total}$$

This is the ratio multiplied by the total count. For example, assume there is only one plot. If observers counted 100 egg nests on this plot, then $M_1=100$. By keeping track of bands, we find there were many instances of reneesting. Assume there were only 50 unique pairs on this plot, the ratio is equal to $50/100 = 0.5$. If the total count is 1000, then the estimated number of breeders is $r*M_{total} = 0.5*1000 = 500$.

The variance estimator is:

$$\text{var}(\# \text{ pairs}) = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^n (y_i - rM_i)^2 \left(\frac{nM_{total}}{N \sum_{i=1}^n M_i} \right)^2$$

We have included an excel spreadsheet that details how to calculate the ratio, the # nests, the variance of # nests, and the 95% CI. The file is: “Estimator 6.xls” and contains equations and citations for adjusting MICs for asynchrony when the entire island is not counted. Estimates computed on the spreadsheet will be accurate as long as the data are accurate and the equations have not been changed.

In practice, bands will have to be used to determine the number of unique pairs, not the number of unique bands. If individuals are monitored on plots, the total number of unique breeding pairs can easily be determined.

- b. Advantages:
- i. This method will provide an estimate of the true number of breeding birds in a given year.
 - ii. Measures of precision are statistically defensible.
 - iii. We expect less variance in estimates of breeding birds and, therefore, statistical power should increase.
 - iv. Counts of breeders are technically easy to derive (see attached spreadsheet). Hence, no special expertise will be required to calculate the number of breeding birds.
 - v. This method will provide much the same information as a large-scale mark-recapture study, but with reduced effort and disturbance.
 - vi. The method will still yield ancillary information that will be useful for understanding the basic ecology of these species and allow the parameterization of population models. For example, ancillary information would include survival rates, breeding propensity, reneesting propensity, etc.
- c. Disadvantages:
- i. While we expect statistical power to increase, there is no pilot data for testing this method. We do not know how large the increase in power will be.

- ii. This method will be more labor intensive and will cause more disturbance on plots.
 - iii. Detectability of nests for MIC methods will still have to be addressed.
- d. Assessment: For species that breed asynchronously and are variable in the number of pairs that nest more than once in a season or renest after failure, we think this is the best approach to investigate. However, there is no pilot data for testing this method. It is promising, because it would facilitate an accurate count of the number of breeding pairs without the investment in data collection and analysis that is required by Alternative 6 (see above).
While this method is statistically defensible, will yield estimates of the numbers of breeding birds that are closer to true numbers, and will yield information required for population models, we do not know how large any gains in statistical power will be. Therefore, we suggest that this method be applied experimentally and evaluated for one or two species.

APPENDIX D. PLOT NUMBERS FOR MONITORING REPRODUCTIVE PARAMETERS

The number of plots that need to be sampled for breeding success and other reproductive parameters is dependent upon these factors: (1) The average number of individuals in each plot (\bar{M}). As the average number of individuals increases, fewer plots need to be sampled. (2) The variance between plots (s_p^2). As the variance between plots decreases, fewer plots need to be sampled. (3) Desired precision. As precision decreases, fewer plots need to be sampled. (4) The total number of possible plots (N). As the total number of plots decreases, fewer plots need to be sampled.

Of these four variables, we have pilot information regarding all factors except the total number of possible plots. This is important, because recommendations for sampling reproductive parameters in the assessment assume 100 plots are available to be sampled. If fewer than 100 plots are available (i.e., if $N < 100$), then fewer plots will need to be sampled than what we recommend. Hence, it is likely that our recommendations in the assessment are conservative (if fewer than 100 plots are available on most islands).

In this supplement, we provide three tables that illustrate how many plots are required for $N = 100$, $N = 75$, and $N = 50$. These tables are based upon the following equations from Scheaffer et al. (1996). The number of plots to be sampled (n) should be estimated as:

$$n = \frac{Ns_p^2}{ND + s_p^2},$$

where N is the total number of possible plots. Other parameters include:

$$D = \frac{B^2\bar{M}^2}{4},$$

where B is the “bound” or desired 95% confidence interval in the proportion (i.e., breeding success) and M is the number of individuals averaged across plots. Finally, s_p^2 is the variance of the proportion, as calculated between plots. This is also described in Section IV of the assessment:

$$s_p^2 = \frac{\sum_{i=1}^n (a_i - \hat{p}m_i)^2}{n-1},$$

where a_i = number of individuals fledged on plot (i), m_i = number of individual eggs on plot (i), and \hat{p} = the proportion of individuals fledged/eggs.

The tables that follow illustrate how the minimum number of plots to be sampled is determined by the variance between plots (s_p^2), the desired level of precision (95% CI), and the average number of individuals in each plot (bold numbers as column headings). Table D.1 assumes that 100 total plots are possible, Table D.2 assumes 75 total plots are available to be sampled, and Table D.3 assumes that 50 total plots are available for sampling. For example, assume that a 95% CI of 0.1 is desired and there are on average 25 individuals in each plot. If the variance between plots is 40, then 20 plots would have to be sampled if $N = 100$, 19 if $N = 75$, and 17 if $N = 50$.

Table D.1. Minimum number of plots to be sampled when 100 possible plots are available for sampling.

Variance between plots	95% CI	Average number of individuals per plot (bold numbers) Minimum number of plots (matrix of numbers)				
		10	25	50	75	100
5	0.01	95	76	44	26	17
	0.05	44	11	3	2	2
	0.10	17	3	2	2	2
	0.15	8	2	2	2	2
10	0.01	98	86	62	42	29
	0.05	62	20	6	3	2
	0.10	29	6	2	2	2
	0.15	15	3	2	2	2
20	0.01	99	93	76	59	44
	0.05	76	34	11	5	3
	0.10	44	11	3	2	2
	0.15	26	5	2	2	2
40	0.01	99	96	86	74	62
	0.05	86	51	20	10	6
	0.10	62	20	6	3	2
	0.15	42	10	3	2	2
60	0.01	100	97	91	81	71
	0.05	91	61	28	15	9
	0.10	71	28	9	4	2
	0.15	52	15	4	2	2

Table D.2. Minimum number of plots to be sampled when 75 possible plots are available for sampling.

Variance between plots	95% CI	Average number of individuals per plot (bold numbers) Minimum number of plots (matrix of numbers)				
		10	25	50	75	100
5	0.01	72	61	39	24	16
	0.05	39	11	3	2	2
	0.10	16	3	2	2	2
	0.15	8	2	2	2	2
10	0.01	74	67	51	37	26
	0.05	51	19	6	3	2
	0.10	26	6	2	2	2
	0.15	14	3	2	2	2
20	0.01	74	71	61	49	39
	0.05	61	30	11	5	3
	0.10	39	11	3	2	2
	0.15	24	5	2	2	2
40	0.01	75	73	67	59	51
	0.05	67	43	19	10	6
	0.10	51	19	6	3	2
	0.15	37	10	3	2	2
60	0.01	75	74	70	64	57
	0.05	70	50	25	14	9
	0.10	57	25	9	4	2
	0.15	44	14	4	2	2

Table D.3. Minimum number of plots to be sampled when 50 possible plots are available for sampling.

Variance between plots	95% CI	Average number of individuals per plot (bold numbers) Minimum number of plots (matrix of numbers)				
		10	25	50	75	100
5	0.01	49	43	31	21	14
	0.05	31	10	3	2	2
	0.10	14	3	2	2	2
	0.15	8	2	2	2	2
10	0.01	49	46	38	29	22
	0.05	38	17	6	3	2
	0.10	22	6	2	2	2
	0.15	13	3	2	2	2
20	0.01	50	48	43	37	31
	0.05	43	25	10	5	3
	0.10	31	10	3	2	2
	0.15	21	5	2	2	2
40	0.01	50	49	46	43	38
	0.05	46	34	17	9	6
	0.10	38	17	6	3	2
	0.15	29	9	3	2	2
60	0.01	50	49	48	45	41
	0.05	48	38	22	13	8
	0.10	41	22	8	4	2
	0.15	34	13	4	2	2