Integrating Physiological and Demographic Parameters into NRDA

A Final Report Submitted to

The Coastal Response Research Center

Submitted by

Dr. Florina S. Tseng
Dr. Victor Apanius
Dr. Ian C.T. Nisbet
Department of Environmental and Population Health
Tufts Cummings School of Veterinary Medicine
200 Westboro Road
West Grafton MA 01536

May 2005 to May 2007

June 5, 2007

Revised October 11, 2007

This project was funded by a grant from NOAA/UNH Coastal Response Research Center.
NOAA Grant Number: NA04NOS4190063. Project Number: 05-959

Abstract

The Natural Resource Damage Assessment (NRDA) process strives to estimate the injury to wildlife populations from oil spills and mitigate human responses to assure recovery of ecosystem services. Estimates of animal population sizes have confidence intervals based on measurement error and the natural temporal and spatial variability of the sampled population. Understanding the magnitude and causes of pre-spill (baseline) variability is crucial for interpreting the demographic injury that is estimated following an oil spill.

The first objective was to characterize the baseline demographic variability in four populations of seabirds in coastal New England, USA, that are typical of wildlife populations that are at risk from oil spills. We studied: Common Terns (*Sterna hirundo*) breeding at Monomoy National Wildlife Refuge (1970 - 1985) and on Bird Island, Marion, MA (1970 - 2002), Roseate Tern (*Sterna dougallii*) breeding on Bird Island (1970 - 2002) and on Falkner Island, Stewart B. McKinney National Wildlife Refuge (1978 - 2002). The Roseate Tern is a federally listed endangered species and the Common Tern has special conservation status in many states. Breeding population size was highly variable, with long-term (decadal) trends accounting for more than half of baseline variability in some cases. Correlations between populations at different sites were due to long-term trends and there is no evidence that inter-annual variation in population size was linked across the region. After accounting for long-term trends, the residual variation indicated that a reduction of 15 to 30% of these tern populations could occur under natural circumstances, based on the 95% prediction interval. Estimates of natural loss rates would be greater if long-term trends were not recognized. The seasonal timing of breeding appeared to be entrained in decadal cycles across the populations and may provide a metric for gauging long-term shifts in environmental conditions or resources used by these species. Measurements of egg production and quality showed the least baseline variability and unnatural perturbations in these parameters appear to be readily detectable. Reproductive success to the point of fledging was the most variable demographic parameter and appeared to respond to long-term population dynamics, regional environment conditions, and local predation pressure. Variation in annual production of fledglings was correlated across sites in the regional ecosystem.

Given the scope of baseline variability of demographic parameters used in NRDA, our second objective was to evaluate the effectiveness of physiological measurements to provide ancillary injury metrics. Twelve hematological parameters were measured in Common Terns on Bird Island in two years prior to, the year of, and two years following an oil spill of opportunity. Hematocrit measurements indicated a higher frequency of anemia following the spill, compared to natural variation. Albumin and uric acid measurements were able to discriminate between favorable and unfavorable natural environmental conditions. Physiological parameters indicated that ecological conditions for the tern population were not unusually unfavorable in the year of the spill. Future NRDAs can employ physiological parameters to provide additional insight into the causes of demographic variation in the aftermath of a spill. The integration of physiological with demographic parameters can more effectively inform decision-makers about wildlife population health and recovery status than the current reliance on demographic responses alone.

Keywords: Seabirds, population variability, oil spill impact, hematology

Acknowledgements

This study would not have been possible without the contributions of Carolyn Mostello (Massachusetts Division of Fisheries and Wildlife), Jeremy Hatch, Michael Brady (Monomoy NWR), Jeffrey Spendelow (USGS), and Michele Sims. The following individuals provided invaluable assistance in the field: Suzanne Conlon, Margaret Friar, Tim Meehan, Chris Buelow, Matt Withroder, Britt Heidinger, Mark Haussmann, Julia Tims, Adam DiNuovo, Michele Kuter, Monica Williams, Anna Ludi and Lee Bar-Sagi. We are grateful for the full cooperation of the Town of Marion. We appreciate the administrative support provided by Carolyn Corsiglia, Betsy Like, Kimberly Newman and Kathy Mandsager. Carol-Ann Manen, Nancy Kinner, and the members of the Scientific Advisory Panel provided valuable advice throughout the project.

This project was funded by a grant from NOAA/UNH Coastal Response Research Center. NOAA Grant Number(s): NA04NOS4190063. Project Number: 05-959

Table of Contents

List of Figures and Tables

1.0 Introduction

In the course of Natural Resource Damage Assessment (NRDA) following the release of crude petroleum or refined products into the environment decisions are made on the degree of injury to wildlife populations and how they should be remedied. The extent of pre-spill information on wildlife populations on which to base these decisions varies dramatically with geography and taxa. Assessments have to consider the size of the pre-spill population of the impacted wildlife and, inevitably, those estimates entail measurement error and the natural variation that occurs as populations are sampled across space and time (Thompson et al. 1998) . A contentious issue in wildlife injury assessments is the magnitude of pre-spill natural (baseline) variation and the extent to which it obscures oil-spill induced demographic changes. A better understanding of the magnitude and causes of baseline demographic variability in wildlife populations likely to be impacted by oil spills is a high priority for the oil spill response community.

Because this natural baseline variation imposes a form of the Heisenberg uncertainty principle, i.e. a phenomenon is assumed to be present but is unobservable, into NRDA that is based solely on demographic parameters, the oil spill community would benefit from the deployment of advanced technology to provide additional metrics of the health of wildlife populations. The measurement of hematological parameters through a combination of on-site methods and overnight express shipment of blood samples to a veterinary clinical laboratory using automated analyzers and rapid internet data distribution has the potential to provide urgently needed insight into the causes of demographic variation following an oil spill.

Considerable progress has been made in understanding the patho-physiological consequences of animals exposed to crude oil and refined products. A review of toxicological dosing studies by Leighton (1993) showed that acute petroleum intoxication leads to: (1) malabsorption secondary to erosion of the gastrointestinal mucosa; (2) hepatocellular dissociation with hemosiderosis; (3) hemolytic anemia with Heinz body formation and reticulocytosis; and (4) renal tubular necrosis and dysfunction. Biomarkers such as Heinz body anemia and liver microsomal (CYP1A) enzyme induction are diagnostic indicators of chronic hydrocarbon exposure and have been associated with depressed growth, reproduction and survival of wildlife (Miller et al. 1978, Leighton et al. 1983, Yamato et al. 1996). Polycyclic aromatic hydrocarbons released from weathered petroleum in sediments were embryotoxic to fish, hepatotoxic to seabirds and slowed the demographic recovery of sea otters for up to 10 years after the Exxon Valdez oil spill (Peterson et al., 2003). Unfortunately, routine measurement of these biomarkers or bioassays during NRDA is currently not feasible. The evaluation of currently commercially available biomarkers for assessing wildlife injury associated with spilled oil is a high priority for the oil spill community.

A number of previous studies have used hematological biomarkers for post-spill assessment of rehabilitated animals and wildlife population health (Newman et al. 2000, Sieser et al. 2000, Golet et al. 2002) but were not integrated into demographic analyses. In order to be effective in the NRDA process, physiological parameters have to be linked to the relevant demographic parameters that are used by decision-makers during spill response and recovery operations.

1.1 **The *Bouchard No. 120* as a Spill of Opportunity**

The *Bouchard No. 120* oil spill of 27 April 2003 in Buzzards Bay, MA, coincided with initiation of breeding by the federally endangered Roseate Tern (*Sterna dougallii*) and the state-listed common tern (*Sterna hirundo*). These widely ranging avian predators integrate prey resources throughout the estuarine ecosystem and are readily accessible in dense nesting colonies, making them ideal study species. Because of their special status, these colonies have been the subject of long-term ecological, demographic and, more recently, hematological investigations.

Terns historically bred at three colonies within Buzzards Bay (Figure 1a). At Bird Island, population and reproductive data have been collected on Common and Roseate Terns since 1970. Ram Island was re-colonized by terns in 1993 and has been monitored continuously since that time. Penikese Island has been continuously monitored since being re-colonized in 1998 by Common Terns and 2003 by Roseate Terns. In addition, population and reproductive data has been collected on Common Terns at Monomoy National Wildlife Refuge (1970-1985) and on Roseate Terns on Falkner Island since 1978 (Figure 2b).

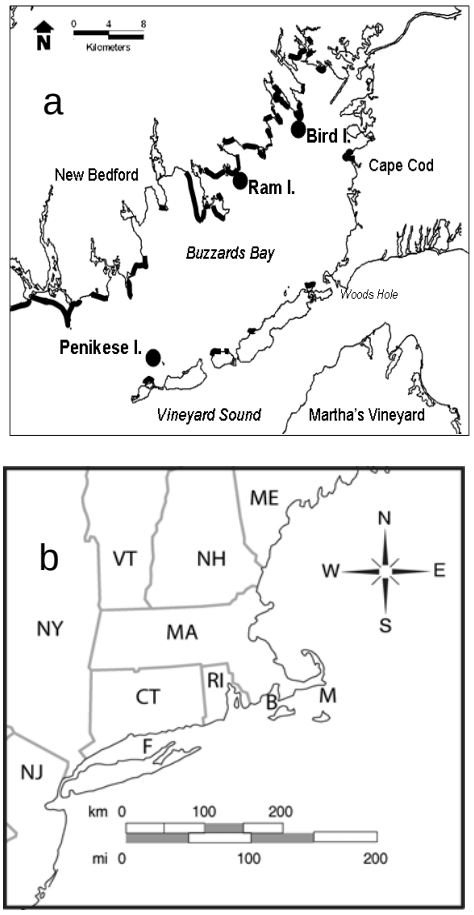


Figure 1. Map of tern colonies in Buzzards Bay, with the extent of oiling in 2003 indicated by thick lines on coastline (a). Location of oiled (B = Bird Island) and reference sites (F = Falkner and M = Monomoy Islands; b).

In the aftermath of the *Bouchard No. 120* spill, avian mortality (mainly loons and seaducks) was greatest in the two weeks following the spill (Figure 2). This date corresponded to the initiation of breeding of Common Terns on Bird Island and the start of blood collection for physiological analyses for studies similar to those we have conducted in previous years. On 5 June 2003, the JAT authorized Entrix Inc. to fund additional blood collection. These samples are not available for analysis due to confidentiality agreements.

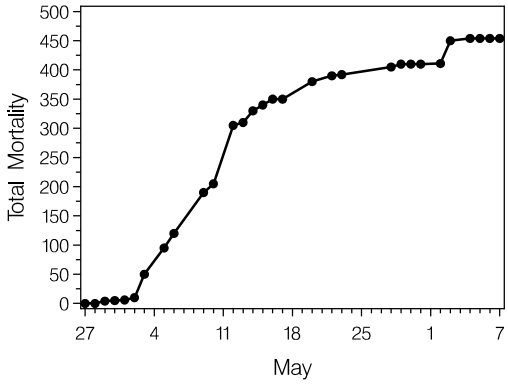


Figure 2. Total number of dead birds recovered in Buzzards Bay after the *Bouchard No. 120* oil spill.

2.0 Objectives

2.1 Baseline Population Variability

The long time series of demographic data for two seabird species at three coastal New England sites provides an unrivalled opportunity to characterize the extent of natural variation in demographic parameters used for wildlife injury assessment. The time-lines for the populations are shown in the upper portion of Figure 3. Demographic data in the year of the oil spill and afterward were excluded to avoid conflict with on-going NRDA of the *Bouchard No. 120.*

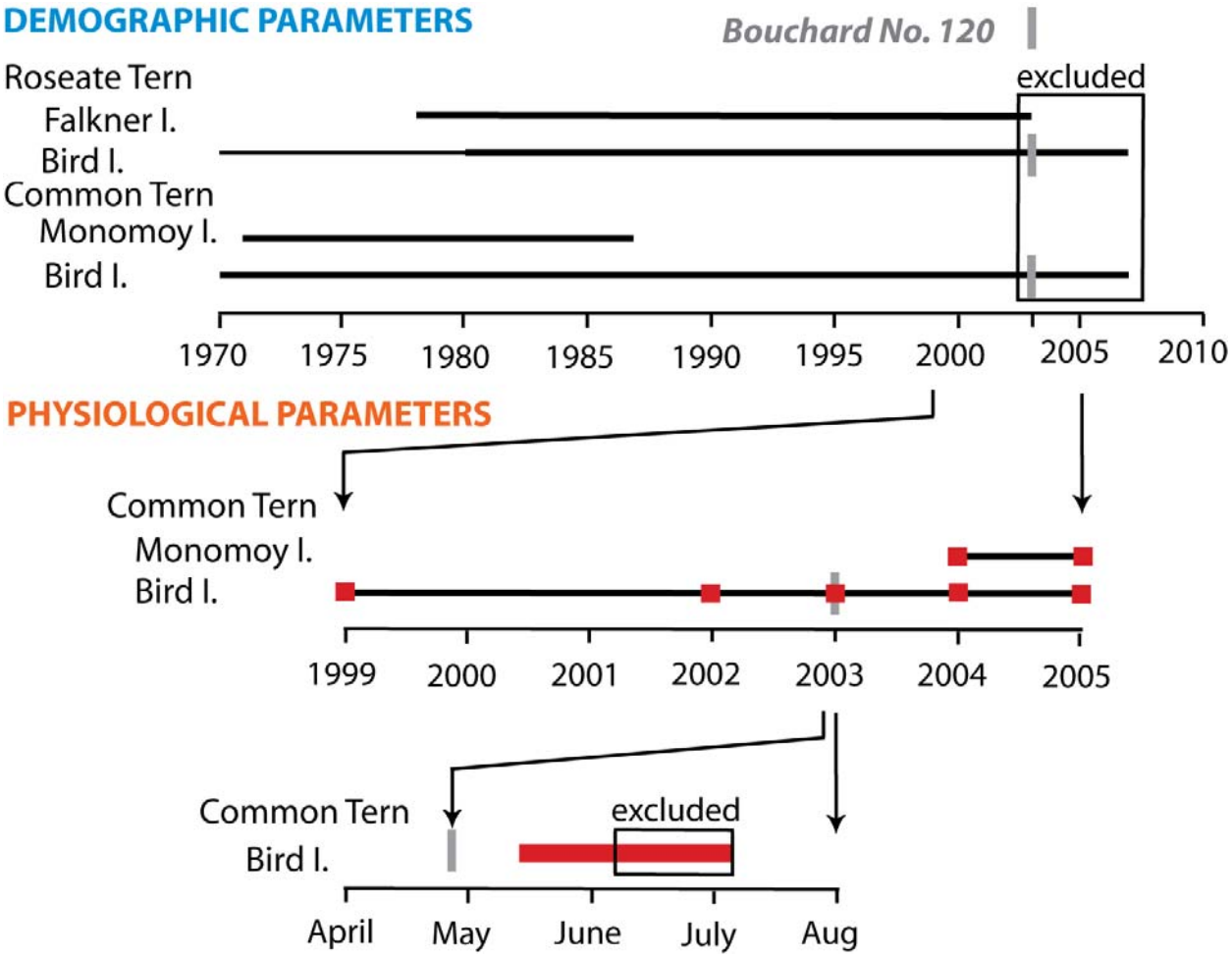


Figure 3. Time line of data collected on the demographic parameters and physiological parameters measured in this study. Under demographic parameters, thick lines denote complete data on all parameters, thin lines show that data on all six parameters were not collected. Gray bar on each timeline denotes the date of the *Bouchard No. 120* oil spill. Red squares/bars indicate years/period of blood sampling.

Demographic data was collected in a standardized manner across populations and we focused on six parameters that relate to population size, seasonal timing of breeding, and reproductive output, in terms of the number and quality of eggs that were laid, and the number of offspring reared to the point of flight (fledging). Our objectives were to characterize natural variability in these parameters by examining the time-frame of variation, i.e. year-to-year, short-and long-term trends, and summarizing the variance in a standardized dimensionless statistical measure- the coefficient of variation (CV). We compared CVs between populations to assess whether they were consistent or population-specific and between the various demographic parameters to identify the most stable vs the most variable. The natural variability that we observed allowed us to estimate the 95 % prediction interval for each parameter, which set an upper limit on the expected deviations that would occur under natural conditions. We used power analysis to estimate the sampling effort, in terms of the number of populations or years, required to detect unnatural large deviations with 70, 80 and 90 % certainty.

2.2 Physiological Parameters

As shown in the lower part of Figure 3, blood samples were collected in pre-spill years 1999 and 2002 as part of our long-term ecological research on Common Terns on Bird Island. In 2003, blood samples were collected shortly after the spill, as terns commenced breeding. We excluded blood samples that had restricted usage due to the on-going NRDA. Post-spill blood samples were collected in 2004 and 2005 from Common Terns on Bird Island and at Monomoy National Wildlife Refuge. The latter site was a reference site in the event that we observed persistent physiological changes at the spill site (Bird Island).

We analyzed a panel of twelve hematological parameters that included indicators of inadequate nutrition, dehydration, overexertion, anemia, and pathological tissue damage in muscle, liver and kidney. We examined these data for shifts in the mean, unequal variances, and the frequency of extreme values in spill vs non-spill years as well as between the non-spill years in order to assess natural variation in these hematological parameters. We were also able to address whether the hematological parameters differed between the two sites in 2004 and 2005, with regard to natural variation or pathological changes.

2.3 Integration of Physiological and Demographic Parameters

Ultimately, the physiological and demographic parameters that we measured serve as proxies for environmental quality and both parameters should co-vary in the same direction in years with contrasting environmental conditions. The demographic analysis revealed that of the four non-spill years with hematological data, two of the years were typical in terms of breeding population size, timing of breeding and reproductive success, while the other years had unusually low, for natural conditions, values for these same parameters. We interpreted these demographic differences to be driven by the ambient environmental conditions and questioned whether this natural environmental variation would be reflected in the physiological parameters. It would be a significant outcome if physiological parameters could resolve between wildlife populations experiencing natural variation in environmental conditions at one level and unnatural pathological changes at another level.

3.0 Methods

3.1 Demographic Parameters

Study Sites and Species. Tern species, the location of breeding colonies and their status as oiled or reference sites are shown in Table 1.

Tern Species	Island / Location	Coordinates
Common	**Bird** Buzzards Bay, 5 km SE of Marion, MA	41° 40' N, 70° 43' W
Common	Monomoy SE corner of Cape Cod, 8 km S of Chatham, MA	41° 32' - 41° 38' N, 69° 58' - 70° 01' W
Roseate	**Bird** Buzzards Bay, 5 km SE of Marion, MA	41° 40' N, 70° 43' W
Roseate	Falkner Long Island Sound, 8 km S of Guilford, CT	41°13'N, 72°39'W

Table 1. Tern species and location information for study sites. Oiled species/sites are shown in bold, reference sites are shown in normal font.

Bird, Ram, and Penikese Islands are located 10-26 km apart in Buzzards Bay, Massachusetts, USA (Figure 1a). All three sites had supported Common Tern colonies in the 1930s and 1940s, but two of them had been overrun by Herring Gulls (*Larus argentatus*) in the 1950s and 1970s and were not reoccupied until the 1990s. The colony with the longest history of continuous occupation is at Bird Island, which had been occupied by Common and Roseate Terns since at least the 1930s and supported 1000-2000 pairs of each species in the 1940s and 1950s (Austin 1944, and unpubl. data). Terns were largely displaced from Bird Island by Herring Gulls in the late 1960s, but increased again after gulls were controlled in the 1970s.

For Common Terns on Bird Island, demographic data have been collected annually since 1970 and we used summary statistics (means or median) through 2002. We also included data on the breeding population size at satellite colonies for a meta-population analysis. For Common Terns on Ram Island, we collected data in 1970 and 1971, the site was deserted in 1972, and then re-colonized, with annual data collection since 1993 and analysis through 2002. For Common Terns on Penikese Island, the colony was re-colonized in 1998, with subsequent annual data collection to the present and analysis through 2002.

For Roseate Terns on Bird Island, data on breeding population size have been collected annually since 1970 and we analyzed data through 2002. Data on median egg-laying date and clutch size have been collected since 1980, and we analyzed data through 2002. Data on clutch and A-egg mass have been collected since 1980, with a gap between 1991 to 1996, and we analyzed data through 2002. Data on productivity have been collected since 1986 and we analyzed data through 2002.

Monomoy National Wildlife Refuge is located 40-60 km east and north of the Buzzards Bay tern populations and on the opposite side of Cape Cod (Figure 1b). Recent studies have shown very

little interchange of Common Terns between this population and those in Buzzards Bay (Tims et al. 2004, and unpubl. data). This site is dynamic, in the sense that its location and conformation changes with erosion and re-deposition. Terns colonized this site around 1961, soon after it was cut off from the mainland and red foxes (*Vulpes vulpes*) were removed. Numbers built up rapidly to 2000-4000 pairs of Common Terns and 200-400 pairs of Roseate Terns in the late 1960s and 1970s, but the Roseate Terns and most of the Common Terns abandoned the site in 1981 and 1985-1987, respectively, following several years of nocturnal predation by Great Horned Owls (*Bubo virginianus*; Nisbet and Welton 1984). Following a gull control program in 1995 and the loss of two other major colony sites in Cape Cod Bay, Common Terns re-colonized South Monomoy Island in 1995 and increased rapidly to over 8000 pairs in 2002. We will refer to this population as Monomoy Island.

For Common Terns on Monomoy Island, demographic data were collected between 1971 and 1987, with occasional missing data on clutch size, clutch mass, and A-egg mass in the early 1980s. The colony was re-established after 1996, but data collection was not started until 1998 and the time series is too short to be useful.

Falkner Island is a unit of the Stewart B. McKinney National Wildlife Refuge. Since the 1960's, Falkner Island has been one of the largest breeding colonies of Roseate Terns (Nisbet and Spendelow 1999).

For Roseate Terns on Falkner Island, data were collected on breeding population size, median-egg-laying date and productivity since 1978, and we analyzed data through 2002. Data on clutch size have been collected since 1980, clutch mass and A-egg mass since 1988, and we analyzed these data through 2002.

The demographic parameters with systematic data collection on the four populations included:

Breeding Population Size. This is the number of nests established during the peak period of nesting (usually the first 25 days of each season) multiplied by two. This represents the most precise estimate of the number of birds at the colony that have the potential to contribute to the local and regional population; nests established after the peak period usually contribute few birds that recruit to the breeding population in subsequent years. This parameter has a central role in injury assessment for seabirds.

Median Egg-laying Date. Egg-laying date is the date that the first egg is laid by each breeding female. The median is used because of the skewness of the distribution of egg-laying dates and is the best measure of central tendency in Common Terns and other seabirds. It represents the population's response to natural (yearly) variation in environmental conditions, with delayed egg-laying in response to low food availability and inclement weather; trends over longer (decadal) time-scales reflect climatic cycles such as the North Atlantic Oscillation. Exposure of seabirds to oil may delay egg-laying directly due to impaired reproductive, endocrine or hematological function. Alternatively, disruption of prey populations by spilled oil may mimic poor food availability under natural conditions and could delay the median egg-laying date of the impacted population.

Mean Clutch Size. Clutch size is the total number of eggs laid by each female and is sensitive to food availability and foraging conditions under natural conditions, with reduced clutch size associated with low foraging success. As with median laying date, mean clutch size could be negatively associated with spilled oil, either through direct contamination and toxicity to the birds or through the negative impact of spilled oil on the population's prey base.

A-egg Mass. This is the mean mass of the first laid egg and, under natural conditions, is sensitive to food availability during the formation of the first egg. Oiled seabirds may lay smaller eggs due to direct and/or indirect effects discussed above.

Clutch Mass. This is the total mass of all the eggs laid by a breeding female and represents an integration of foraging success during the period of egg formation. These data were collected on the most frequent clutch size for each species- 3-egg clutches for Common Terns and 2-egg clutches for Roseate Terns. This parameter may respond to direct or indirect effects of oil spill in parallel to A-egg mass.

Productivity is the average number of offspring fledged by each breeding pair and integrates a number of processes that exhibit natural variability: (1) clutch size; (2) hatching success; and (3) nestling survival. These components vary directly with foraging success of the parents. In addition, hatching success and nestling survival are also influenced by predation, both through direct mortality and through reduced attentiveness by the parents. The combination of these factors can produce substantial variability in productivity. Productivity is an important parameter for predicting population recovery from natural or spill-associated population perturbations.

Our demographic analyses used data prior to the oil spill and that are available in the public domain. We did not analyze demographic data in the year of the oil spill or afterward to avoid conflict of interest with the on-going *Bouchard No. 120* NRDA.

Statistical Analysis. The demographic parameters represent repeated measurements of the same population of individuals at the same sites over multiple years. Each year of data is not statistically independent but is not a strict repeated measures sampling design since individuals were not resampled across time. Time series analyses appeared to be the most appropriate approach to analyzing the characteristics of the demographic data.

We found that our initial time series models, e.g autoregressive integrated moving average (PROC ARIMA, SAS Inc., Cary, NC, USA, 27513, Brocklebank and Dickey 2003.), were not very effective because these models had difficulty simultaneously estimating the autoregressive and moving average components (Figure 4a). This was probably due to non-stationary disturbance, i.e. fluctuations in the strength and direction of perturbations (errors). Autoregressive conditional heterogeneity models can account for time-dependent heteroscedasticity but there was insufficient data for the solutions to converge on reliable parameter estimates. Spline (loess) regression was attempted but also suffered from a shortage of data to determine the number of knots (breakpoints) and it became necessary to input the number of knots. This degree of subjective analysis seemed little better that a *post-hoc* fitting of linear and cubic regression lines.

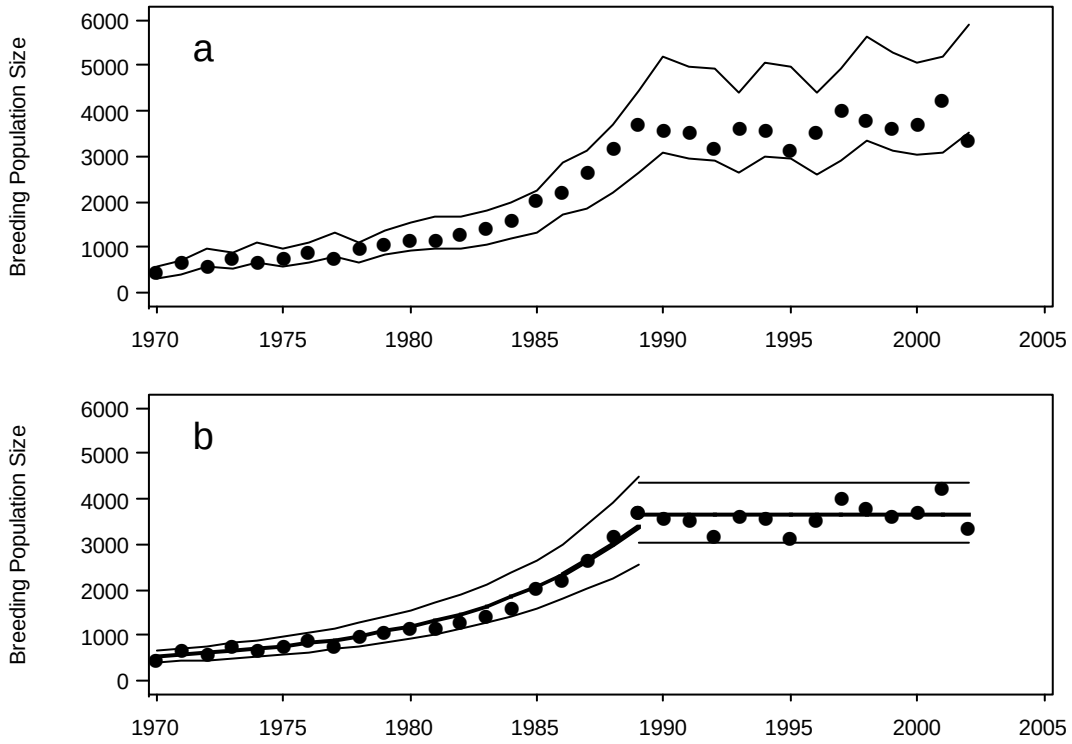


Figure 4. Comparison of time series (a) vs. segmented regression (b) analysis of breeding population size of Common Terns on Bird Island. Thick lines are fitted by regression of log-transformed data. Thin lines are the 95% prediction interval.

Our goal was to determine the minimum prediction interval and not to test hypotheses about population trends. The evidence of non-stationarity indicated that there were complex temporal mean trends and covariances in the data. Therefore, we based model selection on the pattern of residuals from ordinary least squares regressions to determine the number of separate regressions. With this approach, we minimized the prediction interval during periods of low variance and increased the interval based on evidence from the data (Figure 4b).

For exploratory analysis, we fitted polynomial functions to the complete time series to identify short-term trends (PROC GLM, SAS Institute Inc.). As a general rule, the number of years with data was used to specify the degree of polynomial. Time series spanning 5-10 years used a simple first order parameter (YEAR) and were only capable of detecting monotonic trends. Time series spanning 10-20 years used first- and second-order terms (YEAR and $YEAR^2$), which allowed detection of monotonic and concave trends. Time series greater than 20 years used first, second and third order terms (YEAR, $YEAR^2$, and $YEAR^3$), which allowed detection of trends with three phases, e.g. increasing, decreasing, and increasing and vice versa. We matched the order of the polynomial to the sample size in order to be conservative in detecting statistically significant trends and avoid "over-fitting" based on a short run of years. These exploratory analyses indicated inflection points of changing demographic trends which guided the next step of the analysis.

We fit linear or quadratic terms for each time period identified in the exploratory analysis and based the fit on minimization of the residual error variance- the root mean square error. The Durbin-Watson (DW) statistic was used to assess the significance of autocorrelation of the residuals. In most cases, and in agreement with the time series (ARIMA) models, significant levels of autocorrelation were not observed. In the several cases where a significant DW statistic was found, regressions employing the Yule-Walker equations (PROC AUTOREG; SAS Inc) were used to improve the fit and correspondingly reduce the root mean square error, and hence the prediction interval (Zar 1996).

Power analysis was performed for a range of CV values (PROC POWER, SAS Inc.) and the results expressed as graphs showing the proportional change of the mean value that would be significant ($\alpha = 0.05$) at given level of power ($1-\beta = 0.70$, 0.80, or 0.90) as function of sample size (Cohen 1988).

3.2 Hematological Parameters

Field Collection of Blood Samples. In 1999, blood samples were collected on Bird Island between 29 May and 18 June (N=104), and from birds that were captured within 5 days of egg-hatching. Additional blood samples were collected on 26-27 June (N = 14), from birds with young that were within 3 days of fledging. In 2002, blood samples were collected on Bird Island between 16 May and 12 June, although hematocrit measurements were made from birds at the start of incubation (16 May, N=17) or within 3 days of egg-hatching (6-12 June, N = 44). In 2003, blood samples were collected on Bird Island between 17 May and 5 June (N=79), as the birds were incubating eggs. In 2004, blood samples were collected on Bird Island on 10 (N=9) and 12 (N=13) June, and on Monomoy Island on 26 June (N=23). In 2005, blood samples were collected on Bird Island on 16 (N=13) and 17 (N=13) June, and on Monomoy Island on 22 June (N=20). In both 2004 and 2005, and at both sites, birds were captured and blood samples collected within 3 days of egg-hatching. Blood was collected by venipuncture of the jugular or brachial vein and stored on ice in the field. Hematocrit measurements were made between 3-6 hrs after sampling, at which time clotted whole blood was centrifuged at 5,000 rpm for 10 min and serum collected. Serum samples were stored at -20C for a maximum of 30 d and at -80C thereafter.

Hematocrit. Hematocrit is the proportion of whole blood that consists of red blood cells and minor natural variation can be attributed to the animal's hydration status. In pathological states, extreme values of hematocrit are observed as a result of kidney dysfunction (Fudge 2000). Hematocrit is used to diagnose anemia, which is a pathological reduction in the number of circulating red blood cells. Anemia is a hallmark of oil intoxication (Jessup and Leighton 1996, but see Newman et al. 1999) and decreased hematocrit can provide evidence of wildlife injury for petroleum exposure.
Analytical method: Heparinized microhematocrit tubes were filled with 70 microliters (µl) of whole blood and centrifuged at 10,000 rpm for 5 min. The fraction of the total blood column that consisted of packed red blood cells was determined with calipers (±0.1 mm) or with a nomogram. Determinations were made in duplicate.

Total protein. The total protein content of blood plasma shows natural variation attributable to the plane of protein nutrition, with higher total protein levels observed in animals with sufficient dietary protein levels. Inadequate food resources can be signaled by decreased total protein levels with little or no change in hematocrit (Dawson and Bortolotti 1997). In pathological states, total protein levels are markedly decreased in protein-losing enteropathies, such as gut mucosal erosion and kidney glomerular leakage, where plasma proteins are excreted rather than being retained at these barriers (Kaneko 1997). Decreased total protein levels can also signal liver damage in concert with other metabolite and enzyme parameters (Fudge 2000).

Albumin. Albumin is the most abundant protein in blood plasma; it serves as a transporter for many metabolites as well as a source of amino acids for peripheral tissues. Under natural conditions, albumin levels vary with the plane of protein nutrition and high levels are indicative of good body condition in birds (Ghebremeskel et al. 1992). In pathological states, abnormally low values signal liver dysfunction, since this is the site of biosynthesis, as well as protein-losing enteropathies described for total protein (Kaneko 1997).

Uric acid. This metabolite is the end-product of protein digestion in birds and circulating levels of this metabolite depend on the species' diet. In fish-eating (piscivorous) penguins and carnivorous birds of prey, uric acid levels are dramatically increased during digestion of large meals. In natural conditions, uric acid levels in seabirds should reflect the level of food resources, with decreased uric acid levels in birds that have not recently fed. In seabirds, decreased uric acid levels have been associated with poor body condition (Jeffrey et al. 1985, Ghebremeskel et al. 1992, Hollemen et al. 2001, Alonso-Alverez et al. 2002). In pathological states, abnormally high levels of uric acid can signal kidney dysfunction (Fudge 2000).

Blood Urea Nitrogen. This metabolite is a key intermediary linking protein and amino acid metabolism and circulating levels are regulated by the dual action of liver and kidney function. In birds with high-protein diets, blood urea levels are elevated during digestion of large meals. As with uric acid, blood urea nitrogen levels reflect short-term nutritional status of individual animals and the availability of food resources. In pathological states, abnormally elevated levels of blood urea nitrogen are indicative of impaired kidney function or extreme dehydration (Fudge 2000).

Triglycerides. These metabolites are the ingested and mobilized forms of dietary lipids and elevated levels indicate a high plane of nutrition in animals with a lipid-rich diet, such as seabirds. Under natural conditions, triglyceride levels vary with food availability, especially high quality foods with a high lipid-density. In pathological states, extreme values of triglycerides indicate the inability of the liver to regulate lipid metabolism.

Cholesterol. The levels of this metabolite in seabirds, which typically ingest a low-cholesterol diet, should be insensitive to food resource levels, but circulating levels may be elevated and reflect mobilization during starvation states. Cholesterol levels were observed to be elevated in birds in poor body condition (Jeffrey et al. 1985, Ghebremeskel et al. 1992). In pathological states, extreme values of cholesterol would indicate the liver's inability to maintain homeostatic levels.

Bile acids. These metabolites facilitate absorption of ingested lipids and natural variation is expected to mirror triglyceride levels, with higher levels reflecting increased food availability. In pathological states, abnormally high levels would indicate impaired liver function due to leakage of bile acids into circulation.

Creatine kinase. This enzyme is released from skeletal muscle during strenuous physical activity (Fudge 2000; Young and Bermes, 2001) and would be elevated in naturally adverse conditions, where inclement weather or limited food availability induces greater foraging effort. In pathological states, creatine kinase is released by a variety of tissues, principally by damaged muscle tissues, but not by the liver (Fudge 2000).

Lactate dehydrogense. This enzyme is released from skeletal muscle during strenuous physical activity (Fudge 2000; Young and Bermes, 2001) and would be elevated in naturally adverse conditions, where inclement weather or limited food availability induces greater foraging effort. LDH activity was consistently elevated in seabirds from oiled habitats 9 and 10 years after the Exxon Valdez oil spill (Golet et al. 2002). In pathological states, lactate dehydrogenase is released by a variety of damaged tissues and the source is indicated by the pattern of changes of additional enzyme parameters. Elevated levels of lactate dehydrogenase indicate liver damage if creatine kinase levels are normal (Fudge 2000).

Aspartate aminotransferase. This enzyme is released from skeletal muscle during strenuous physical activity (Fudge 2000, Young and Bermes, 2001) and would be elevated in naturally adverse conditions, where inclement weather or limited food availability induces greater foraging effort. In pathological states, aspartate aminotransferase is released by a variety of damaged tissues and, in concert with other enzymes, can indicate impaired liver function. Elevated levels of aspartate aminotransferase indicate liver damage when lactate dehydrogenase levels are also elevated, but creatine kinase levels are normal (Fudge 2000). Elevated aspartate aminotransferase levels were observed in seabirds from oiled habitats but not reference sites 8 and 10 yrs after the Exxon Valdez oil spill (Seiser et al. 2000, Golet et al. 2002). Elevated aspartate aminotransferase and creatine kinase levels, with normal lactate dehydrogenase levels have been reported in oiled and rehabilitated seabirds (Newman et al. 2000).

Metabolite, protein and enzyme parameters were measured with methods described in the Appendix (Pages 70-79). Assay reactions were measured with a 96-well microplate reader (Victor3, PerkinElmer, Shelton, CT, 06484). Each assay was performed in triplicate. Human normal control serum (Cat. No. 1902-050, Thermo-Electron, Inc., Louisville, CO, USA, 80027) was used as a reference standard that was analyzed in triplicate at the beginning and end of each run of eight serum samples

Statistical Methods. The hematological parameters that we measured have the potential to distinguish between favorable and adverse natural conditions, as well as between natural and oil-spill associated variation in three ways. First, the variance of the parameter can be increased as a result of the loss of physiological homeostasis due to stressful ecological conditions. Differences in the variances between populations were tested using Levene's F-ratio test. Second, a consistent increase or decrease of the parameter between populations may be evident through a shift of the mean value. Differences of means between populations were tested using the

omnibus F-ratio from an analysis of variance (ANOVA) or, when the populations being compared had unequal variances, using Welch's F-ratio test. Third, some individuals may show extreme values of the parameter that represent pathological states, even in the non-spill populations. Depending upon the number of individuals and the magnitude of the deviation, this would potentially be detectable as a shift in the means and or a greater variance in the population with abnormal values. Alternatively, extreme values can be detected by constructing 95% confidence intervals of values from a reference population. Following standard practice in clinical hematology, we present the frequency distributions of the hematological data in order to assess these three criteria for identifying physiological differences between populations (Solberg 2001, p 254). Examination of the distributions ensures that assumptions of parametric statistical tests, e.g. continuity, equal variances, and gaussian distribution of errors, are not violated. Many physiological parameters have asymmetrical or skewed distributions, with a greater spread of the high values compared to the low values, which can normalized by logarithmic transformation (Solberg 2001, p 254). This was the case for total protein, uric acid, triglycerides, and the enzyme analytes. For these analytes, data shown in tables and figures are the untransformed values, whereas statistical tests were performed after log-transformation. Following the omnibus test, multiple comparisons tests were used to identify which population means significantly differed from each other. To adjust the P-values for these multiple comparisons tests we used randomization (bootstrap) methods that are robust to unequal sample sizes and variances, and distributional assumptions (Westfall and Young 1993; Westfall et al. 1999).

We examined the potential effect of date on the hematological parameters in two ways. First, multiple individuals were typically sampled on the same date and those individuals may have been influenced by local weather or environmental conditions on that date. Consequently, we used ANOVA with sampling date as a categorical variable to test the significance of this source of heterogeneity. Second, seasonal variation in environmental conditions may have produced monotonic trends in the physiological parameters. In this case, we used sampling date as a continuous variable to identify temporal trends in the physiological parameters.

4.0 Results

4.1 Baseline Variability

We have focused our analysis on demographic traits that: (1) would be most relevant to injury assessment; (2) have the most complete time series; and (3) are comparable across species and sites. For all four populations, we have comparable data on the following parameters.

Breeding population size The analysis of breeding population size of the four populations clearly showed the dynamic nature of tern colonies in coastal New England (Figure A1). Common Terns appeared to have a greater extent of population variability, with populations changing by an order of magnitude, several hundreds to several thousands. Population variability in Roseate Terns was within an order of magnitude. 1,000 to 3,000 on Bird Island or 100 to 400 on Falkner Island.

For both species, the primary source of variation was the long-term trends in breeding population size, which spanned one or two decades, and were apparent in each population. Population sizes

were deemed constant, but highly variable, for only two subsets of data- Common Terns on Bird Island between 1989 to 2002, and for Roseate Terns on Falkner Island between 1978 to 1995. The evidence for an increasing breeding population size for Common Terns on Monomoy Island was marginal (P = 0.075, Table A1). In 5 of the other 8 data subsets, the coefficients of determination (r^2) for models that included temporal trends, e.g. linear or quadratic year effects, ranged from 0.534 to 0.959 (Table A1). This can be interpreted as 53 to 96 % of the variation in breeding population size could be attributed to long-term temporal trends. Estimating the slope of these long-term trends allowed a narrowing of the prediction confidence interval (Figure 4). For example, the prediction interval for the breeding population size of Roseate Terns on Bird Island in 1995 was 1332 - 4651 without accounting for the long-term trend, vs 1620-3490 from the regression model with a time trend, which is 44 % reduction the width of the interval.

Within data subsets exhibiting long-term periods of population increase, decrease or stability, year-to-year variation was the primary source of variation as evidenced by the minor extent of autocorrelation (Table A1). The autocorrelation coefficient was significant in two data subsets and the more complex models that accounted for the autocorrelation did little to reduce the residual error about the regression line (Table A1). Thus, the population size each year appeared to be statistically independent events, with no evidence of "memory" of the preceding year.

The historical decline of the Common Tern population at Monomoy occurred in the same years that the Bird Island population was increasing, which induced a strong negative correlation between these two populations (r = -0.739, N = 17, P = 0.009; Table A2). The only other significant correlation between population size was for Roseate Terns at Bird and Falkner Islands (r= 0.654, N = 25, P = 0.0004; Table A2). Deviations from the long-term trends, i.e. residuals from the regression lines or means, were not significantly correlated, indicating that annual fluctuations in population size were not synchronized across any of the populations, species or sites (Table A2).

Except for the decline stage of the Common Tern population at Monomoy Island (CV=7.35%, Table A1), the CVs for breeding population size ranged between 0.98 and 2.51. Including all Common Tern populations, e.g. Bird, Ram, and Penikese Islands, into a single meta-population (1995-2002) yielded a highly significant regression model with a CV= 1.11%. The maximum year-to-year reduction that could be attributable to natural variation ranged from 16% for the population with minimum population variability (Common Tern, Bird Island, 1989-2002) to 35% for the most variable population (Roseate Tern, Falkner Island, 1978-1990). Detecting losses outside this range would require a modest number, e.g. 2 to 3, years of sampling effort, based on the power curves (Figure A2).

Median egg-laying date. Across the 1970s, the median-egg-laying dates decreased monotonically in the two populations with data, Common Terns at Monomoy and Bird Islands (Figure A3, Table A3). During the 1980s, median egg-laying dates increased in all populations. After the early 1990s, the median egg-laying dates started decreasing again, with the possible exception of Common Terns on Bird Island (Figure A3, Table A3). The long-term synchronization in egg-laying date is also evident from the high correlation coefficients among three of the populations (r= 0. 589 to 679); the correlation between Common Terns on Monomoy

and Bird Islands was not significant (Table A4). It appeared that this parameter was responding to regional climatic variation.

The long-term temporal trends had r^2 values between 0.379 and 0.852, implying that a substantial proportion of baseline variability was accounted by decade-long trends. Autocorrelation was significant in two data subsets, but modeling this covariance yielded only marginal improvement in the corresponding CVs: 14.0 vs 15.8 % for Common Terns, Bird Island, 1980-1992, and 10.2 vs 10.3 % for Roseate Terns, Falkner Island, 1978-1992 (Table A3).

With CVs ranging between 6.7 and 15.8 %, a corresponding shift in egg-laying dates by greater the 15 to 30% would be considered unusual with respect to natural variation. Given this level of natural variability, the detection of an anomalous delay in egg-laying would require 5 to 10 years of sampling effort (Figure A2).

Mean clutch size. For Common Terns at Monomoy Island, mean clutch size increased in the 1970s and decreased in the 1980s (Figure A4, Table A5). For Common Terns on Bird Island, clutch size was constant through the mid-1980's, showed a downward trend, then was stable, but highly variable through the 1990's (Figure A4, Table A5). The downward trends in mean clutch size for Common Terns were correlated between Bird and Monomoy Islands ($r = 0.617$, $N = 15$, $P = 0.014$). For Roseate Terns, mean clutch size was stable and less variable across the entire study. Thus, long-term temporal trends contributed to baseline variability in only a minority of the study years. Autocorrelation of mean clutch sizes was negligible (Table A5) and mean clutch sizes of Common Terns were correlated between Monomoy and Bird Islands (Table A6. Common Terns at Bird Island had the lowest and highest CVs (1.66 and 8.10%), which corresponded to a 3.5 to 18.4 % decrease in clutch size attributable to natural variation. The detection of anomalous changes in clutch size would require between 3 to 10 years of sampling effort (Figure A2).

A-Egg mass. Temporal trends in the mean mass of the first laid egg were significant in all populations, except Roseate Terns on Falkner Island, with r^2 values ranging between 0.315 to 0.987, although the later value may be artifactually high since it was based on 4 collinear data points (Figure A5, Table A7). For Common Terns at Bird Island, there was a significant degree of autocorrelation, which after taken in account, reduced the statistical significance of the temporal trend ($P = 0.0066$ to 0.044), but did little to reduced the residual error (Table A7). A-egg masses were not correlated among populations; however, the residual value was weakly correlated between Common and Roseate Terns on Bird Island ($r = 0.548$, $N = 14$, $P = 0.042$; Table A8), suggesting that local environmental conditions were influencing deviations in A-egg masses from the long-term averages.

The CVs ranged between 0.28 and 1.07 %, although the minimum value was also based on the four collinear data points (Table A7). Using the CV for the population with the most robust sample size (1.0%, Common Tern, Bird Island, 1981-2002), the mass of the A-egg would have to be reduced 2.2% in order to exceed natural variation, which could be easily detected after 3 years of sampling (Figure A2).

Clutch Mass. The patterns for clutch mass were essentially identical to A-egg mass, except for slightly greater range of CVs, 0.60 to 1.69 % (Figure A6, Table A9). There was no significant autocorrelation for clutch mass data (Table A8). The correlations between population were not significant, with either the raw data or residuals (Table A10).

Productivity. The number of offspring fledged per breeding attempt showed considerable variability in Common Terns and less, but still substantial variability compared to other demographic parameters in Roseate Terns (Figure A7, Table A11). Productivity varied erratically in Common Terns at Monomoy Island, although a quadratic (concave down) function could be fit before productivity essentially went to zero in 1981. For Common Terns at Bird Island, productivity increased through the mid-1980s, declined, and then appeared to be stable but highly variable. In contrast, Roseate Terns at Bird Island showed stable productivity across years, while those at Falkner Island appeared to follow a shallow decline. Productivity was correlated between Roseate Terns on Bird and Falkner Islands ($r = 0.547$, $N = 18$, $P = 0.019$) as well as between Common and Roseate Terns on Bird Island ($r = 0.701$, $N = 18$, $P = 0.0012$). These correlations were virtually identical for the original or residual values (Table A12). Surprisingly, productivity was also correlated between Common Terns on Bird Island and Roseate Terns at Falkner Island, with a stronger correlation using the original data ($r = 0.781$, $N = 25$, $P < 0.0001$) compared to the residuals from long term trends ($r = 0.478$, $N = 25$, $P=0.016$; Table 11). These correlations across species, populations, and sites, suggests that local, as well as regional environmental conditions may have been entraining reproductive output of these populations. However, local factors such as predation are known to have influenced both trends and year-to-year fluctuations, especially the declines in productivity at Monomoy and Falkner Island (unpubl. data).

Excluding Common Terns on Monomoy Island as representing an exceptional case (resulting from sustained predation pressure), the CVs for Common Terns ranged from 6.6 to 35.9 % on Bird Island, while the CVs for Roseate Terns were 11.9 and 20.9 % on Bird and Falkner Islands, respectively (Table A11). Using the extreme CV values for Common Terns on Bird Island, a reduction in productivity between 13.6 and 66% would be considered within the range of natural variation, while the corresponding reductions are 26.0 and 50.2% for Roseate Terns on Bird and Falkner Islands, respectively. For Common Terns, detection of anomalous changes in productivity could be observed in as little as 3 years or might require as many as 20 years of sampling effort, depending on the CV for that particular population (Figure A2).

Correlations among Demographic Parameters. If demographic parameters are highly correlated ($r > 0.90$), then a single parameter can serve as a proxy and reduce the cost of demographic data collection if the correlations are generalizable across populations. However, the correlation structure of the demographic parameters may be species-, site- or population-specific and each parameters can equally contribute to decisions regarding response, recovery and damage assessment.

For Common Terns on Monomoy Island, population sizes and median egg-laying dates were negatively correlated ($r = -0.764$, $N = 15$. $P = 0.0009$) and mean clutch sizes were positively correlated with mean clutch masses ($r = 0.690$, $N=11$, $P = 0.019$; Table A13).

For Common Terns on Bird Island, population sizes were not correlated with median-egg laying dates but were negatively correlated with all of the reproductive parameters (P = 0.055 to < 0.0001; Table A14). Median egg-laying date was negatively correlated with clutch size (r = -0.402, N = 33, P = 0.021) and productivity (r = -0.482, N = 33, P = 0.0045). Reproductive parameters were inter-correlated (Table A14), notably mean clutch size and productivity (r = 0.917, N = 33, P <0.0001).

For Roseate Terns on Bird Island, the residual population size was negatively correlated with residual laying data (r = 0.699, N = 20, P = 0.0006), while A-egg mass, clutch mass, and productivity were strongly intercorrelated (r = 0.804 to 0.838; Table A15).

For Roseate Terns on Falkner Island, breeding population sizes were positively correlated with productivity (r= 0.538, N = 25, P = 0.0055) and residual egg laying dates were negatively correlated with residual productivity (r = -0.664, N = 25, P = 0.0003; Table A16). Of the reproductive parameters, the correlation between A-egg masses and clutch masses was the strongest (r = 0.819, N = 15, P 0.0002).

The correlations among demographic parameters are consistent with the interpretation that environmental conditions are causing the two parameters to co-vary in the expected direction. However, the correlation structure differed between populations, suggesting that there are not universal relationships between the demographic parameters.

4.2 Hematological Parameters

Hematocrit. One of the key signs of intoxication by crude oil or refined products is anemia and is measured by hematocrit (packed cell volume), which is the proportion of blood occupied by red blood cells by volume. We analyzed hematocrit in concert with body mass variation because both appear to be related to environmental conditions in non-spill years.

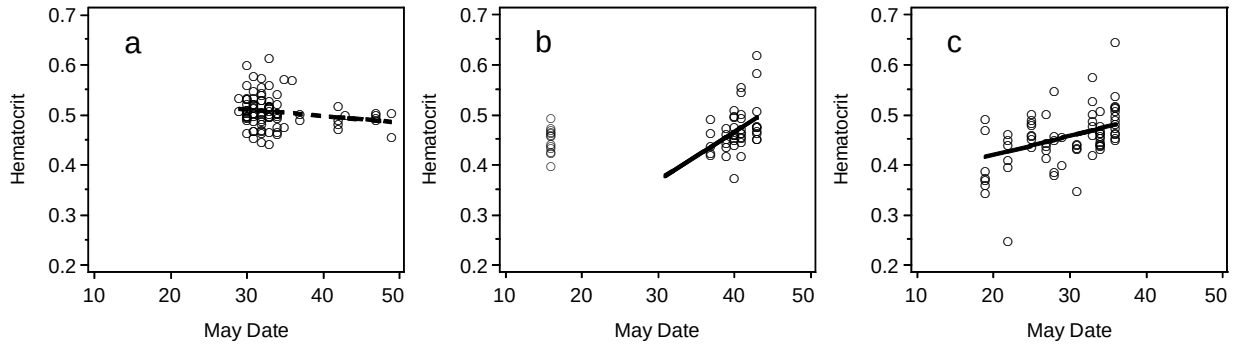


Figure 5. Relationship between hematocrit and calender date, expressed as the number of days after 1 May, for Common Terns on Bird Island in 1999 (a), 2002 (b), and 2003 (c). Solid lines shows significant linear trends (P<0.05), broken lines show marginally significant trends (0.05 < P < 0.10).

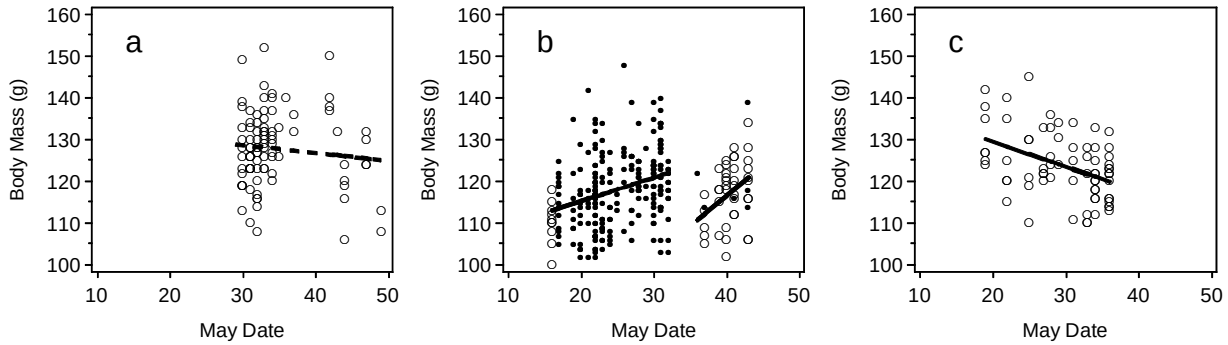


Figure 6. Relationship between body mass and calender date, expressed as the number of days after 1 May, for Common Terns on Bird Island in 1999 (a), 2002 (b), and 2003 (c). In 2002 (b), open circles represent body masses of birds with hematocrit measurements, as in 1999 (a) and 2003 (b). Lines as shown in Figure 5.

In 1999, hematocrit did not vary with the date of sampling ($F_{12,79}$ = 0387, P = 0.58, r^2 = 0.12) but showed evidence of a marginally significant negative trend across dates ($F_{1,90}$ = 3.43, P = 0.067, r^2 = 0.04, Figure 5a). Body mass varied across sampling dates as a consequence of heterogeneity between days ($F_{15,90}$ = 3.65, P < 0.0001, r^2 = 0.38) and only marginally as a monotonic negative trend ($F_{1,104}$ = 3.86, P = 0.052, r^2 = 0.04; Figure 5a). This was due to the re-capture of birds at the time of fledging. Body mass of breeding terns decrease approximately 5% between the time of egg-hatching and fledging, as reported elsewhere (Apanius and Nisbet 2006). Although both hematocrit and body mass showed marginally significant declines across the sampling period, hematocrit was not related to body mass ($F_{1,81}$ = 0.48, P = 0.49, r^2 = 0.01; Figure 7a).

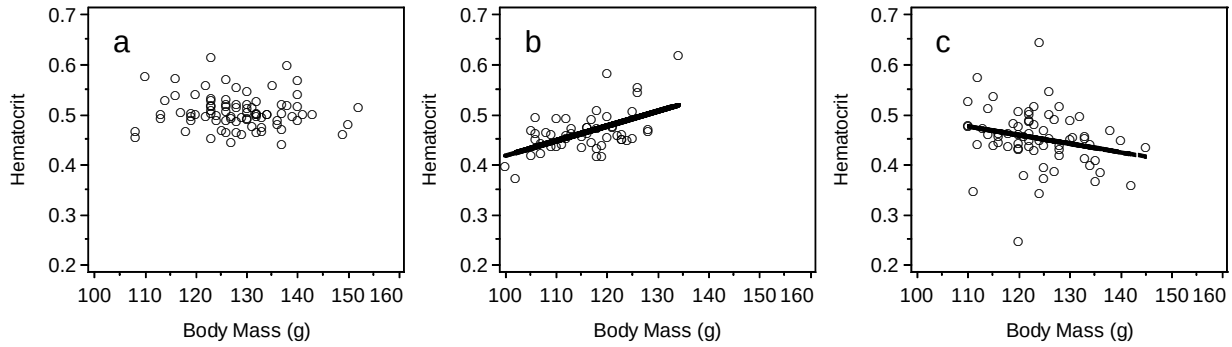


Figure 7. Relationship between hematocrit and body mass of Common Terns on Bird Island in 1999 (a), 2002 (b), and 2003 (c).

In 2002, hematocrit values varied between sampling days ($F_{5,52}$ = 2.81, P = 0.025, r^2 = 0.21), with a significant monotonic increase between days 37 to 42, ($F_{1,39}$ = 6.96, P = 0.012, r^2 = 0.19; Figure 5b). For the dates on which hematocrit was measured (Figure 5b, open symbols), similar results were obtained for the relationship between body mass and date, including the increasing monotonic trend ($F_{1,41}$ = 6.56, P = 0.014, r^2 = 0.16; Figure 6b). The increase in hematocrit and body mass across this short time period was reflected in the significant positive relationship between hematocrit and body mass ($F_{1,52}$ = 19.52, P < 0.0001, r^2 = 0.27, Figure 7b), which may have been due to a common environmental factor. Hematocrit measurements were made after major storm events at the beginning of the breeding season (12-15 May) and around the time of

hatching (7 June). With the limited number of storm events and blood sampling dates, the relationships of hematocrit and body mass with weather variables would be inconclusive. We return to the possible relationships between hematocrit, body mass and weather in a latter section. Across the entire incubation period, body mass showed a significant increase with date (Figure 6b, closed symbols; $F_{1,283} = 10.74$, $P < 0.0012$, $r^2 = 0.036$).

After the *Bouchard No 120* spill on 27 April 2003, hematocrit showed a sustained monotonic increase with date ($F_{1,73} = 16.88$, $P < 0.0001$, $r^2 = 0.19$; Figure 5c). In contrast to results from 2002, body mass in 2003 significantly decreased across dates (Figure 6c; $F_{1,67} = 14.10$, $P = 0.0004$, $r^2 = 0.17$). Also in contrast with 2002, hematocrit in 2003 was weakly and negatively related to body mass (Figure 7c; $F_{1,67} = 3.94$, $P = 0.051$, $r^2 = 0.06$). Analysis of data from 1999, 2002, and 2003showed a highly significant year x body-mass interaction ($F_{1,118} = 15.87$, $P = 0.0001$), which indicated that the relationship between body mass and hematocrit was clearly different before and after the oil spill. In 2003, 25 of 66 (37.9 %) Common Terns were visually scored as having spots or smudges on their plumage that may have indicated contact with petroleum. The presence of these spots or smudges was not related to the bird's hematocrit value ($F_{1,64} = 0.46$, $P = 0.50$, $r^2 = 0.01$) or anemia status ($\chi^2 = 0.09$, df = 1, $P = 0.76$).

In 2004, hematocrit did not differ between the two sampling dates on Bird Island ($F_{1,18} = 2.68$, $P = 0.12$, $r^2 = 0.13$) and the means and variances were comparable between Bird and Monomoy Islands (Table A17). Consequently, data for the two sites were pooled for further analyses.

In 2005, hematocrit was significantly lower on 16 June compared to 17 June (0.454 vs. 0.481; $F_{1,21} = 7.01$, $P = 0.015$, $r^2 = 0.06$), possibly due to high winds on the former date. The variance was significantly lower on the one sampling date at Monomoy Island, but the mean values were not significantly different between Bird and Monomoy Islands (Table A17). Data from the two sites was pooled for the analysis of inter-annual variation.

Figure 8. Distribution of hematocrit values by year. Line at top connect years whose means are not significantly different, i.e. not significantly different. The horizontal reference line is the lower 95% confidence interval for 2002 and 2005 (see text).

The variance in hematocrit was significantly greater in 2003, than in pre- and post-spill years (Levene's $F_{4,304} = 4.41$, $P = 0.0018$) and mean hematocrit was significantly higher in 1999 and 2002 (Figure 8; Welch's $F_{4,134.2} = 24.75$, $P < 0.0001$). There were more birds with unusually low hematocrit values in 2003 than in non-spill years. For 2002 and 2005, the 95% confidence interval for hematocrit in these years of low, but natural, levels was 0.417-0.519. Using the lower boundary of this confidence interval to classify birds as anemic or normal, 15 of 60 (20.0%) birds were

anemic in 2003 compared to 5 of 93 (5.1%) in naturally adverse years and 8 of 259 (2.9%) birds in all non-spill years. The frequency of anemic birds was significantly greater after the spill compared to naturally adverse conditions (relative risk= 4.65, 1.6- 13.5 95% CI, $\chi^2 = 9.22$, df=1, P = 0.0024) or all non-spill years (relative risk = 8.41, 3.41 - 20.17 95% CI, $\chi^2 = 28.29$, df=1, P < 0.0001).

In 2003, the variation in hematocrit and body mass did not appear to be determined by weather conditions, as appeared to the case in non spill years. In non-spill years, hematocrit increased with temperature as measured by the departure from the seasonal norm ($F_{1,230} = 32.11$, P <0.0001, $r^2 = 0.123$, Figure 9a), while there was a marginally significant negative trend in 2003 ($F_{1,73} = 3.33$, P = 0.072, $r^2 = 0.043$, Figure 9b). Thus, the relationship between hematocrit and ambient temperature was clearly different between spill and non-spill years, as indicated by the highly significant year x temperature-departure interaction ($F_{1,303} = 13.46$, P = 0.0003). Body mass variation was not influenced by temperature departure (analyses not shown).

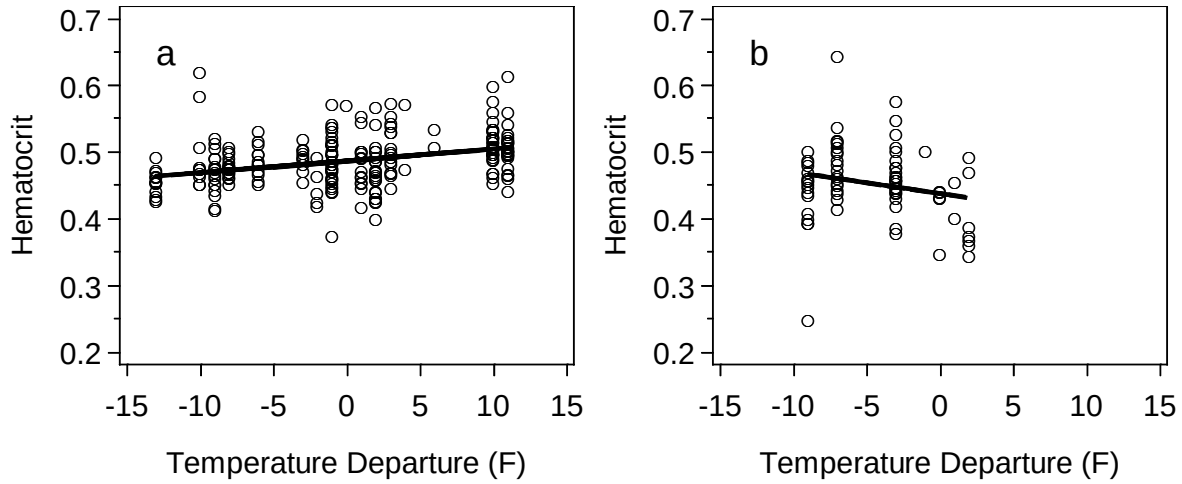


Figure 9. Relationship between hematocrit and temperature departure from seasonal norm, as measured at TF Green Airport NOAA weather station, for Common Terns on Bird Island for non- (a) and oil-spill (b) years.

In non-spill years, hematocrit decreased as average wind speed increased ($F_{1,230} = 22.35$, P <0.0001, $r^2 = 0.088$, Figure 10a), while there was no relationship in 2003 ($F_{1,73} = 0.00$, P = 0.97, $r^2 = 0.000$, Figure 10b). Thus, the relationship between hematocrit and average wind speed was also clearly different between spill and non-spill years, as indicated by the significant year x temperature-departure interaction ($F_{1,303} = 6.26$, P = 0.013). Body mass variation was not influenced by average wind speed (analyses not shown).

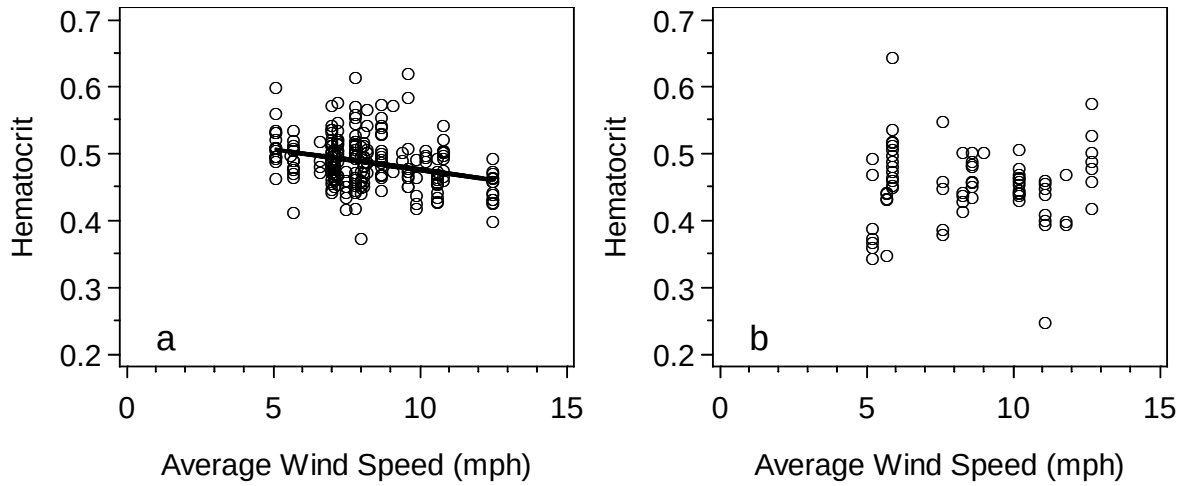


Figure 10. Relationship between hematocrit and average wind speed, as measured at TF Green Airport NOAA weather station, for Common Terns at Bird Island in non- (a) and oil-spill (b) years.

In summary, hematocrit showed seasonal variation within years under natural conditions and this variation appeared to be linked to flying or foraging conditions as influenced by temperature and wind speed. In years with naturally adverse environmental conditions, as evidenced by delayed breeding and low reproductive output, hematocrit reflected these ecological conditions. After the oil-spill, 20% of the hematocrit values were unusually low, compared to natural variation, and the mean hematocrit increased with date after the spill. This variation could not be explained by variation in body mass or weather conditions in that year.

Total Protein. Total protein means and variances did not differ between Common Tern populations on Bird and Monomoy Islands in either 2004 or 2005; accordingly, sites were pooled in the comparison between years. Total protein concentration had the greatest variance and mean in 1999 (Table A17, Figure A8). Mean total protein concentration was lower in 2002, and the lowest means were observed in 2004 and 2005. In the oil-spill year, mean total protein concentration was intermediate between the non-spill years with naturally high and low levels, but was significantly lower than the highest mean observed in 1999.

Albumin. Albumin concentration did not differ between Common Tern populations on Bird and Monomoy Islands in either 2004 or 2005 and sites were pooled for each year. The variances in albumin concentration did not differ between years (Table A17). Mean albumin concentration was significantly lower in years with naturally adverse environmental conditions, e.g. 2002 and 2005, compared to more favorable years, e.g. 1999 and 2004, respectively (Figure A9). There appeared to be a downward shift in mean albumin concentration across the non-spill years. Mean albumin concentration in the year of the oil-spill was comparable to the non-spill year 1999 (Figure A9). However, neither a significant shift in the mean, increased variance, nor increased frequency of extreme values was observed in the year of the oil spill.

Uric Acid. Uric acid means and variances did not differ between Common Tern populations at Bird and Monomoy Islands in either 2004 or 2005; sites were pooled. The variances in uric acid concentrations did not differ between years (Table A17). The distributions were non-normal and

appeared bimodal (Figure A10), justifying the use of the distribution-free (bootstrap) multiple comparisons of means that we used. Mean uric acid concentration was highest in 1999 and 2004 and significantly lower in 2002 and 2005. Uric acid concentration in the year of the oil spill was intermediate and not significantly different from the years with high levels (1999 and 2004) and those with the low levels (2002 and 2005; Figure A10).

Blood Urea Nitrogen. The variances in blood urea nitrogen concentrations did not differ between Common Tern populations at Bird and Monomoy Islands in either 2004 or 2005, but the mean concentration was significantly higher in Common Terns on Bird Island in 2005. We pooled the data from Bird and Monomoy Island in 2004, but retained two separate groups in 2005. The variances in blood urea nitrogen concentration did not differ between years (Table A 17). Mean blood urea nitrogen concentration was greatest in 1999 and significantly lower in all other years, including the oil-spill year. The exception was Bird-Island in 2004, which was intermediate and not significantly different from any other year (Figure A11).

Triglycerides. The variances of triglyceride concentration did not differ between Common Tern populations at Bird and Monomoy Islands in either 2004 or 2005 (Table A1), but the mean concentration was significantly lower in Common Terns on Monomoy Island in 2005. We pooled the data from Bird and Monomoy Island in 2004, but retained two separate groups in 2005. The variances in triglyceride concentrations did not differ between years (Table A17). Mean triglyceride concentration was greatest in 1999 and significantly lower in other non-spill years, except in comparison to Bird Island in 2004 (Figure A12). In the year of the oil-spill, mean triglyceride concentration was intermediate and not significantly different between the years with high levels (1999 and Bird-Island in 2005) and years with low values (2002, 2004, and Monomoy Island in 2005). In the context of the larger number of populations being compared and controlling the significance level for the multiple comparisons, the difference in mean triglyceride concentration between Bird and Monomoy Islands in 2005 was not significant in the larger analysis (Figure A12). Repeating the analysis with the two island populations combined in 2005, showed essentially the same pattern: significantly higher levels in 1999 compared to other non-spill years, with the values in the oil-spill year being intermediate.

Cholesterol. The variances in cholesterol concentrations did not differ between Common Tern populations at Bird and Monomoy Islands in either 2004 or 2005, but the mean concentration was significantly higher in Common Terns on Monomoy Island in 2005. We pooled the data from Bird and Monomoy Island in 2004, but retained two separate groups in 2005. The variances in cholesterol concentrations did not differ between years (Table A17). Mean cholesterol concentrations were significantly lower in 2004 compared to other non-spill years, with the exception of an intermediate mean value for Bird-Island in 2005 (Figure A13). In the oil-spill year, mean cholesterol concentration was intermediate and not significantly different between the non-spill years with high levels (1999, 2002 and Monomoy-Island in 2004) and the years with low levels (2002 and Bird Island in 2005). It is noteworthy that this parameter did not distinguish between contrasting natural conditions between 1999 and 2002 yet showed significant variation between two sites in one of the two years.

Bile Acids. The means and variances of bile acid concentration did not differ between Common Tern populations at Bird and Monomoy Islands in 2004 or 2005 (Table A1). Accordingly, sites

were pooled in the comparison between years. Preliminary analysis revealed little variation in bile acid concentration between spill and non-spill years and, consequently, very few samples (N=6) were analyzed in 2002 when sample volumes were limiting. Excluding data from that year, the means and variances in bile acids concentration did not differ between years (Table A17, Figure A14).

Lactate Dehydrogenase. The variances of lactate dehydrogenase activity did not differ between Common Tern populations at Bird and Monomoy Islands in 2004 or 2005, but the mean activity was significantly higher in Common Terns on Bird Island in 2005. We pooled the data from Bird and Monomoy Island in 2004, but retained two separate groups in 2005. The mean and variance of lactate dehydrogenase activity was greatest in 1999 and lowest in 2002, with intermediate values in other years, including the oil-spill year (Table A17, Figure A15).

Creatine Kinase. The variances of creatine kinase activity did not differ between Common Tern populations at Bird and Monomoy Islands in 2004 and 2005 (Table A1), but the mean activity was higher in Common Terns on Bird Island in 2005, but with marginal statistical significance (Table A17). Nonetheless, we pooled the data from Bird and Monomoy Islands in 2004, but retained two separate groups in 2005. The means were significantly lower and variances were significantly greater in 2003 and 2004, compared to 2002 and 2005/Bird-Island. (Table A17, Figure A16). However, the pattern of variation in creatine kinase activity was not clearly discernable.

Aspartate Aminotransferase The variances of aspartate aminotransferase activity did differ between Common Tern populations at Bird and Monomoy Islands in 2005, but not in 2004 . The mean activities did not differ between Bird and Monomoy Islands in 2005, after controlling for the unequal variances. However, the mean activities differed between Bird and Monomoy Islands in 2004. Therefore, we pooled the data from Bird and Monomoy Islands in 2004, but retained two separate groups in 2005. The variances of aspartate aminotransferase activity did differ between years (Table A17). Mean aspartate aminotransferase activities did significantly differ between years (Figure A17).

4.3 Integration of Demographic and Physiological Parameters

The key innovation of this project is the ability to use a long-term demographic study to provide an environmental context for the physiological parameters that were measured in the non-spill years. From the relative ranking of demographic parameters (Table 1), 2002 and 2005 were years with poor reproductive outcomes for Common Terns on Bird Island, with delayed median egg-laying date, small clutch sizes, low clutch masses, and low productivity. Curiously, A-egg mass did not fit the pattern, which was otherwise consistent with environmental conditions that were unfavorable for reproduction. In contrast, 1999 and 2004 appeared to be relatively typical years in terms of reproductive outcomes. The ranks of breeding population size did not appear to be coupled with this marked inter-annual variation in reproductive success. Nonetheless, we feel justified by the reproductive data to regard two years (1999 and 2004) as having normal and the other years (2002 and 2005) as having unfavorable natural conditions. A more detailed analysis using the original data, as opposed to ranking by summary statistics, will be conducted upon release of confidential data in 2008.

Parameter	N	1999	2002	2004	2005
Breeding Population Size	16	10	3	8	11
Median Egg-laying Date	16	14.5	4.5	7.5	1
Mean Clutch Size	16	8	1	9.5	2
A-Egg Mass	16	7.5	14	6	1
Clutch Mass	9	4	3	7	NA
Productivity	16	7	2	13	1

Table 2. Rank of demographic parameters of Common Terns on Bird Island between 1989 and 2005, excluding the year of the oil spill. N is the numbers of years with data. Low ranks indicate unfavorable outcomes in terms of reproductive success.

The distinction between years in demographic parameters, principally reproductive output, was clearly evident in the significant downward shift in mean hematocrit between favorable and unfavorable years (Figure 8). The only birds with hematocrit values indicative of anemia in non-spill years, were observed in the unfavorable years, e.g. 2002 and 2005. The relationship between weather conditions and hematocrit under natural conditions further strengthens the interpretation that hematocrit is sensitive to natural variation in environmental conditions. Hematocrit did not vary between the two sites in 2004 and 2005, suggesting the environmental regime that influenced hematocrit was on the regional scale. The critical feature of the hematocrit parameter in the year of the oil spill was the significantly increased variance and a significantly higher frequency of abnormally low values than in years with unfavorable environmental conditions.

One of the plasma metabolites that indicated recent feeding activity, uric acid, also showed significant downward shifts of the mean in years of unfavorable environmental condition, compared to years with favorable ones (Figure A10). The plasma protein that indicated long-term plane of nutrition, albumin, also showed similar significant shifts in mean levels (Figure A9). In the year of the spill, the two parameters were of intermediate (uric acid) or high (albumin) levels, suggesting that environmental conditions were not excessively unfavorable.

Of the remaining physiological parameters, the relationship with demographic parameters was somewhat ambiguous or patently uninformative. The metabolites that indicate recent feeding activity, blood urea nitrogen and triglycerides, showed significant downward shift of mean values between 1999 and 2002, but not between 2004 and 2005; in fact, they indicated that the latter years were both unfavorable. In addition, the means of these two parameters both differed between sites in 2005. The other indicator of long-term nutrition, total protein, showed a similar pattern as blood urea nitrogen and triglycerides, but without the difference between sites. These additional parameters did not suggest that environmental conditions in the year of the spill were exceptional. Cholesterol, bile acids, and the enzyme parameters showed patterns that were not consistent with the previously discussed ones and did not provide additional information on the relationship with demographic parameters. However, they did show that there were not

exceptional changes in tissue function in breeding common terns in the year of the spill, in comparison to the natural variation that was observed before and after the spill.

5.0 Discussion and Importance to Oil Spill Response/Restoration

5.1 Demographic Analysis

A critical facet of NRDA is estimating the extent of wildlife injury due to an oil spill; this requires the ability to distinguish between natural demographic variation and the spill associated perturbation. Our analysis of baseline variability in four seabird populations has identified the magnitude and sources of variation in several key demographic parameters ranging from breeding population size to timing of reproduction to breeding productivity. We address each topic separately.

Breeding Population Size A key feature of NRDA is the prediction of the population sizes of impacted wildlife species in the absence of oil-induced perturbations. These predictions can be based on population sizes before a spill or by comparison with non-impacted reference sites during and after a spill. Our analysis of demographic data allows us to address the question- Are predictions based on a limited number of years consistent with those made with much longer, and expensively collected, time series data?

Our analysis of population size data for four seabird populations revealed that the majority of baseline population variability stemmed from long-term temporal trends that typically spanned a decade or two. When these long-term trends were taken into account in regression models, the remaining "unexplained variance" was further probed. If the errors from the regression models in successive years were correlated, i.e. autocorrelated, then extensive time-series data would produce a more refined prediction of population size. We found that the extent of autocorrelation in the population size data was not consistent and, when present, did not substantially improve the prediction of population size through a reduction of the unexplained variance, or root mean square error in statistical parlance. Therefore, we conclude that the year-to-year variations of increasing, stable, or decreasing populations that we studied were, in general, statistically independent. This is relevant to NRDA because the precision of the population size estimates is constrained by the prediction interval from a regression model, which accounts for the number of years (sample size) and natural baseline variability across those years (variance). These prediction intervals would be unreliable to the extent that the data were non-independent and autocorrelated. Fortunately, these complications were not apparent in the data we analyzed, but if they were present, statistical models could take that into account with sufficient sample sizes. The simple conclusion is during NRDA the assessment of natural variability of pre-spill or non-impacted populations will depend on the extent of data collection efforts, i.e. the number of years or populations that were measured.

The unique aspect of our study was the ability to estimate baseline variability without the constraint of limited sample sizes. This allowed us to narrow the prediction interval by accounting for long-term trends when present. Therefore it is important to point out that estimates of baseline variability of population size, as summarized by the CV, are the "best case" scenarios. With that caveat, we estimated that losses between 16 to 35 % of a breeding

population of terns could be within the range of natural inter-annual variation. The estimates of the sampling effort needed to document losses outside this envelope are also optimistic because long-term trends are taken into account. The root ecological causes for this variation is not known, but the similarity among the four populations and across three decades provides support for the generality of this finding.

An often untested assumption in NRDA is that impacted and reference sites co-vary in response to ecological variation, so that favorable or adverse environmental conditions are similar across sites. We had the opportunity to test that assumption by looking at the correlation between population sizes of the same species at two different sites. If the tern populations responded to favorable environmental conditions by increasing the number of birds breeding and unfavorable conditions by reducing the number, then positive correlations would have been observed. If breeding birds were moving between breeding colonies between years then negative correlations would have been observed between sites. We refined our ability to detect these correlations by also analyzing the residuals from the long-term trends or means, which provide a measure of the departure from overall trend or average. This would be pertinent if long-term trends were obscuring correlations. The residuals could show that the deviations of the demographic parameter were in the same direction each year, regardless of the direction of long-term trends, and provide evidence for a common environmental factor driving demographic variation on regional scale.

For Common Terns, breeding population sizes were negatively correlated between Monomoy and Bird Islands and probably reflected local population dynamics as the correlation between residuals was not significant. This supports banding data that shows that movement of individuals between sites on an annual basis was minimal. For Roseate Terns, breeding population sizes, but not residuals, were positively correlated between Bird and Falkner Island, which may be due to a common response to long-term (decadal) environmental fluctuations. In summary, we could not provide strong evidence that breeding population sizes were linked, either by meta-population movement or via a common response to inter-annual environmental variation. This may reflect the fact that most terns breed every year, without the extensive non-breeding in unfavorable years that characterizes some other seabird species. Regardless of the causal basis for this finding, it suggests that seabird populations in the same region are varying independently under natural conditions.

Phenology. The annual initiation of reproduction was measured using the median egg-laying date, which is sensitive to environmental, principally weather, conditions. All four populations appeared to be entrained in decade-long trends of increasing or decreasing egg-laying dates. The North Atlantic Oscillation was considered as a possibility for driving these cycles but simple correlation analyses were unsuccessful at linking the two phenomenon. Median egg-laying dates and breeding population sizes were negatively correlated for the Common Tern population on Monomoy Island, but not for other populations. For Roseate Terns on Bird Island, egg-laying dates were negatively correlated with breeding population size, using the residuals to remove the long-term trends. Thus, despite the apparent synchronization of breeding phenology across populations, egg-laying date was related to breeding population size, but not in a consistent manner. Although this parameter showed considerable baseline variation, this easily measured parameter has the potential to serve as a proxy of natural variation in environmental conditions.

Reproductive Parameters. The reproductive parameters we measured fall into three categories: egg production, as measured by clutch size; egg quality, as measured by the mass of the first-laid (A-) egg and of the entire clutch of eggs; and production of fledglings, which represents the stage of reproduction that can be conveniently measured. Common Terns were distinguished from Roseate Terns in the variability of clutch size. Both Common Tern populations showed periods of temporal trends and periods with high and low natural variability, while Roseate Tern clutch sizes were stable across two decades, with consistent CVs. For Common Terns, it is not known whether the correlated temporal trends in clutch size between Monomoy and Bird Island were linked through regional environmental conditions. For Common Terns on Bird Island, the strong negative relationship between breeding population size and clutch may represent density-dependent effects, as this growing population was reached carrying capacity around 1988. Clutch size and breeding population size were unrelated in other populations. Clutch sizes were negatively related to the median egg-laying dates in Common Terns, but not Roseate Terns, suggesting that these two parameters varied in tandem with natural environmental conditions in one, but not the other, species. While this parameter had relatively low levels of baseline variability, which would facilitate detection of unnatural perturbations, the general lack of correlation among most populations suggests that comparisons made with reference sites might be tenuous.

Within populations, egg quality measurements, e.g. A-egg mass and clutch mass, were highly correlated with each other and also with clutch size. This suggests that these three parameters convergently assess female body condition, which ultimately reflects natural environmental variation. However, egg quality traits were not consistently related to egg-laying date or breeding population size. As with clutch size, among population correlations of egg quality measures were not significant, although the data available for comparisons were relatively sparse.

A key reproductive parameter for NRDA is the ability of an impacted population to return to pre-perturbation levels, which depends on the species' recruitment rate. We did not measure the number of offspring that returned to the breeding population due to the 2-4 yr pre-reproductive period of these species, which is typical for seabirds. Our measurement of annual fecundity (productivity) was the number of offspring fledged per breeding (nesting) attempt. Productivity was the most variable of the demographic parameters we measured. For Common Terns, long-term temporal trends accounted for approximately half of the variance, with year-to-year variation accounting for the other half. With the exception of Common Terns at Monomoy Island, the inter-annual variation in productivity was highly correlated between populations even between different species at distant sites, e.g. Common Terns on Bird Island vs Roseate Terns on Falkner Island. This strongly suggests that productivity represents an integrated measure of resource availability as well as environmental conditions across regional ecosystems. In relation to breeding population size, productivity was positively correlated (Common Terns on Monomoy Island, Roseate Terns on Falkner Island), negatively correlated (Common Terns on Bird Island), or uncorrelated (Roseate Terns on Bird Island). Productivity was negatively related to median egg-laying date in one population (Common Terns on Bird Island), but not others. As expected, productivity was correlated with egg production and quality, although the strengths of the relationships varied between populations. Baseline variability in productivity was difficult to

characterize because of differences between species, with Common Terns displaying a wide range of natural loss scenarios and Roseate Terns more stable. While this parameter may be difficult to measure and subject to confounding by predation and other natural factors, our observation of a tight linkage between sites offers optimism for its information content.

5.2 Physiological Analysis

The pervasive natural variability of wildlife populations challenges the accuracy of injury assessments during NRDAs. Because the measured extent of wildlife mortality can fall within the envelope of natural variation, additional analytical methods can be employed to gauge the impact of spilled oil on the health of wildlife populations. One of the most common and widely used approaches in wildlife health assessment is blood collection and clinical laboratory analysis. Our study provided the unique opportunity to measure hematological parameters and assess their utility before, during, and after an oil spill of opportunity. We found that measurement of hematocrit, as well as plasma metabolites, proteins, and enzymes, were informative metrics that can be utilized by decision-makers during oil-spill response/recovery operations and in the NRDA process.

Hematocrit This was the most informative parameter for distinguishing between natural variation in environmental conditions and potential pathological consequences of exposure to spilled oil. The mean value of hematocrit shifted downward in naturally adverse conditions as evidenced by the robust comparison of 1999 and 2002, where samples were collected across a range of dates and weather conditions. Our detailed analysis using weather summary data, showed that environmental conditions can influence hematocrit over a relatively short time period. The mean hematocrit is shifted downward during periods of cold and windy conditions associated with spring storm events, which presumably prevents effective foraging. Mean hematocrit shifts upward during warm and calm weather conditions, which may represent hemoconcentration and dehydration during vigorous foraging activity. In the naturally adverse conditions of 2002, hematocrit was positively related with body mass, supporting our interpretation that this physiological parameter reflects prevailing ecological conditions.

After the oil-spill in 2003, mean hematocrit increased with date as if it was responding to improving environmental conditions, but body mass and summary weather data conflicted with this interpretation. Bird body masses were decreasing as the season progressed and mean hematocrit was unrelated to weather conditions. It appeared that hematocrit was decoupled from prevailing ecological conditions and this finding is supported by the unprecedented frequency of abnormally low hematocrit values, indicative of anemia. We used the hematocrit measurements from two naturally adverse breeding seasons to construct reference intervals for identifying atypical hematocrit values in individual birds. We used the stringent criteria of a 95% confidence interval and used data from naturally adverse conditions. The cut-off would have been higher if the reference values were based on "good" and "bad" years combined, but the statistically significant conclusion is unchanged. There was a greater frequency of breeding Common Terns with anemia in the aftermath of an oil spill than would be expected from natural variation in this physiological parameter.

The identification of anemia via abnormal hematocrit as a key physiological metric is significant for the oil spill response community. It has a solid empirical basis from previous experimental and mechanistic studies of petroleum intoxication (Leighton, et al. 1983, Jessup and Leighton, 1996). The assay is rapid, simple and can be performed on site under field conditions, allowing measurements to be made while captured birds are temporally held. It is routinely employed at wildlife rehabilitation facilities that are part of spill response efforts. Species-specific reference values are available in the literature for a wide diversity of wildlife species. Our results indicate that measurements are similar in two populations within the same regional ecosystem, e.g. Bird and Monomoy Islands. Finally, the interpretation of hematocrit values is re-enforced by two other readily available sources of information that are routinely collected- body mass of the individual animal and local weather conditions at the site.

Metabolite Parameters. As with hematocrit, a shift in the mean concentrations of uric acid and triglycerides distinguished between naturally favorable or adverse environmental conditions. This outcome accords well with seabird physiology. Like other piscivorous seabirds, Common Terns have a diet with relatively high proportions of protein and lipids. Following ingestion, plasma triglyceride levels increase as dietary lipids are absorbed and mobilized to the liver and storage (adipose) tissue. The end-product of protein digestion in birds is uric acid and large increases in plasma uric acid concentration, and to a lesser extent blood urea nitrogen, are observed after ingestion of a large meal in piscivorous and carnivorous birds. The observed bimodal distribution of uric acid and, possibly triglycerides, is consistent with birds being in either absorptive, i.e digesting food, or post-absorptive, i.e. fasting, states. The significant shift in means of both parameters between the naturally favorable (1999) and adverse (2002) years is consistent with the expected frequency of encountering birds in fed or fasted states in the two years. It appeared that these two plasma metabolites served as a proxy for short-term food resource availability either through variation in foraging conditions and/or prey vulnerability. Because these parameters indicate foraging success it is important to have robust sample sizes, approximately 40-60 per population, distributed across multiple sampling dates. The difference in triglyceride concentrations between sites in 2005 provides cautionary evidence against inferences drawn from one or two collecting dates per site.

After the oil spill, the mean concentrations of uric acid and triglycerides were within the range observed under natural conditions and the absence of a significant proportion of extreme values indicates that liver and kidney functions were not disrupted in the population studied. An additional end-product of protein digestion, blood urea nitrogen, was measured and the results were similar to uric acid. Two physiological indicators of liver function, cholesterol and bile acids, were measured but were uninformative with regard to resolving natural ecological variation or spill-associated perturbations. Cholesterol levels did not differ between rehabilitated oiled birds and birds from a reference site (Newman et al. 2000).

Measurement of plasma uric acid and triglyceride concentrations provides oil-spill responders with two important sources of information. First, these physiological parameters address whether post-spill demographic changes in a wildlife population are driven by natural ecological processes. Naturally adverse environmental conditions would be signaled by an increased proportion of unfed individuals and a downward shift of means values. Second, these parameters address whether the impacted wildlife populations are directly affected by the patho-

physiological consequences of petroleum intoxication, evidenced by abnormal values, or indirectly through disruption of food resources or foraging habitat.

Protein Parameters. Total protein and albumin concentrations provided information on the long-term nutritional plane of individual animals. Albumin appeared to provide greater resolution between years with naturally favorable (1999) and adverse (2002) environmental conditions. Measurement of albumin concentration provides additional information for decision-makers about the ecological conditions that may be influencing wildlife populations after an oil-spill. Although we did not observe atypical values in the year of the oil spill, measurement of albumin concentration, in concert with cholesterol and bile acids, can be used to assess liver function in animals exposed to oil.

Enzyme Parameters. Measurements of two enzymes, creatine kinase and lactate dehydrogenase, had the potential to reveal muscle fatigue and overexertion in animals foraging in adverse environmental conditions in 2002, but the results were non-significant or contradictory. Creatine kinase was elevated in 2002, but only significantly in relation to 2003 and 2004. Lactate dehydrogenase was significantly elevated in the naturally favorable year, 1999, than in other years. Our assumption that this species would have a greater foraging effort in a year with poor reproductive success may be faulty. The nutritional indicators point to short- and long-term shortfalls in foraging and perhaps flight activity was reduced as an energy-saving strategy. The observation that the mean enzyme levels differed between sites in some non-spill years but not consistently, e.g. creatine kinase in 2005 but not in 2004, suggests additional ecological factors may be influencing these physiological parameters. Measurements of the enzymes had the potential to reveal pathological tissue damage but abnormally high values were not observed in the year of the spill, or in any year for that matter. While these results were less informative than those from the protein and metabolite parameters, the robust ability of these parameters to detect pathological tissue damage in general warrants their inclusion in future work. The measurement of these plasma enzymes permits decision makers to draw stronger inferences with regard to distinguishing between natural eco-physiological variation and oil-associated patho-physiological changes.

5.3 Integration of Physiological and Demographic Analyses

Pre-spill linkage of physiological and demographic parameters. The central premise for the integration of demographic and physiological parameters is that natural variation in environmental conditions will modify an individual animal's physiological status, which, in turn, will affect its reproductive output and, ultimately, will be manifested as a demographic response. This linkage is an active area of basic ecological research. The long-term demographic dataset from Common Terns on Bird Island, together with our recent physiological studies, provides a unique context for addressing the connections between demographic and physiological processes that are ultimately driven by environmental factors.

The four non-spill years in which physiological parameters were measured could be unambiguously dichotomized into two groups based on the demographic parameters. Low breeding population size, late egg-laying dates, small clutch, and low productivity characterized the 2002 and 2005 breeding seasons, while these parameters had intermediate values in 1999 and

2004. While we considered 1999 and 2004 to represent relatively favorable environmental conditions and 2002 and 2005 to represent unfavorable conditions, we should point out that the contrast is between typical conditions and relatively poor conditions, because 2002 and 2005 actually represented the lowest years, in terms of clutch size and productivity since 1988, whereas 1999 and 2004 values were typical for that same span of time. Therefore, the long-term demographic data provides the context to state that the test of the resolving power of the physiological parameters was between two sets of years that were inferred to have typical vs naturally adverse environmental conditions. Without this demographic context, the conclusions drawn from the physiological parameters would have been weaker.

The means of three of the physiological parameters, hematocrit, albumin and uric acid, were significantly different between favorable and unfavorable years. Additional parameters, triglycerides, blood urea nitrogen and total protein, provided supportive evidence for distinguishing between natural environmental conditions but were not as unambiguous as the former three parameters. The physiological role and direction of change of these parameters is consistent with their connection to demographic parameters via ambient environmental conditions. We showed that hematocrit varied with weather conditions across non-spill years in order to discount the possibility that the occurrence of abnormally low hematocrit values in the year of the oil spill was due to natural causes. The unanticipated relationship between hematocrit and local weather under natural conditions points to the utility of incorporating local weather and climatic information with physiological parameters. Therefore, we emphasize the importance of collecting and analyzing a wide range of environmental and ecological data at impacted and reference sites. Because the physiological indicators are linked to proximal environmental conditions, restoration scientists and decision makers should be aware that the physiological parameters are more sensitive to local ecological conditions than demographic parameters.

Post-spill physiological responses. The public record showed that two dead Common Terns were recovered in the 2003 *Bouchard No. 120* spill event. Thus it appeared at first sight that the demographic impact of the oil spill on the Common Tern population was minimal. Yet our analysis of hematocrit suggests that an estimated 20% of the breeding Common Terns showed signs of anemia that could not be attributed to natural causes and was consistent with adverse health effects from oil exposure. Thus, we were able to use physiological parameters to identify sub-lethal injury to wildlife resources in the aftermath of the spill.

The fact that post-spill anemic birds were not underweight suggests that anemia was an acute response to exposure, possibly from preening oiled plumage. From a clinical point of view, hematocrits that are below the normal reference values (typically 80% quintile cut-offs based on the precautionary principle) are associated with reduced aerobic stamina; this would be expected to have resulted in reduced flight capacity and hence reduced foraging success in terns, which have high energy demands at the time of egg-laying. Since we do not have repeated measurements from the same individuals, we do not know whether the post-spill, seasonal increase in hematocrits in 2003 represented recovery within individuals. We can not dismiss the possibility that anemic birds had reduced survival and were lost from the population.

Our measurement of physiological parameters in an impacted wildlife population after an oil spill is similar to previous studies. The typical approach is to measure physiological parameters,

e.g. body mass, blood metabolites/ enzymes, etc, in a non-impacted population to establish species-specific reference values for these parameters. These reference values are then used to gauge whether the parameter values from the impacted population are abnormal and represent pathological states consistent with hydrocarbon intoxication, e.g. liver and kidney dysfunction. A central assumption of this approach is that the physiological reference values are stable properties of the species and do not vary with ecological conditions. Our demonstration that three physiological parameters respond to ecological variation has important consequences for the incorporation of physiological parameters in NRDA. First, inferences about pathological changes in wildlife after an oil spill using reference values from historical or distant populations of the same species will have to be validated for the impacted and reference populations.

Another critical assumption is that the established reference values are invariant with respect to sample handling, analytical methods, and reference standards. A central tenet of veterinary clinical chemistry is that each laboratory should establish their own reference values for each species and particular breed or population in order to compare patient values. Our study is unique because we compare the values of physiological parameters measured in the same laboratory and in the same wildlife population before, during, and after an oil spill. Therefore, we feel confident that, apart from hematocrit, pathological tissue changes were not observed in breeding Common Terns following the oil spill. Thus, we believe that the anemia we observed represented a transient response to preening oiled plumage and ingestion of oil sufficient to induce liver or kidney damage did not occur in breeding Common Terns after the spill.

This section demonstrates both the potential utility and the potential pitfalls of incorporating physiological parameters in post-spill response and recovery operations as well as in NRDA. The ability to assess the health status of surviving individuals in an impacted population provides a metric for calibrating post-spill demographic changes. As we have shown, natural baseline variability in demographic parameters can obscure significant losses in population size and reproduction. While it can be argued that these natural fluctuations are evidence of the species' demographic resiliency, it assumes that the surviving individuals are healthy and physiologically fit. It also assumes the habitat and ecosystem services have been restored to provide sufficient food resources for the recovering population. Thus, physiological parameters provide decision-makers with information on two fronts.

First, demographic losses due to oil exposure would be signaled by pathological changes in surviving individuals and serve to indicate that a proportion of the population that was exposed to the deleterious effects of oil. We do not expect a simple relationship between mortality and morbidity due to oil, but the presence of birds showing signs of oil intoxication shows that wildlife injury is present in the absence of extraordinary (>15-30 %) demographic losses. Secondly, the ability of physiological parameters to track post-spill ecological conditions can inform decision-makers whether chronic oil exposure is occurring and whether the wildlife populations are limited by recovery of food resources. The panel of physiological parameters in our research project can provide information in these two areas and potentially provide crucial information to improve the cost-effectiveness of post-spill mitigation efforts.

6.0 Technology Transfer

A key innovation of our research was the integration of physiological with demographic processes, We have shown how existing technology in veterinary clinical hematology can be utilized for informing decision makers responsible for post-spill responses and restoration. We envision that physiological parameters can be integrated into demographic analyses of wildlife injury in the progression illustrated in Figure 11.

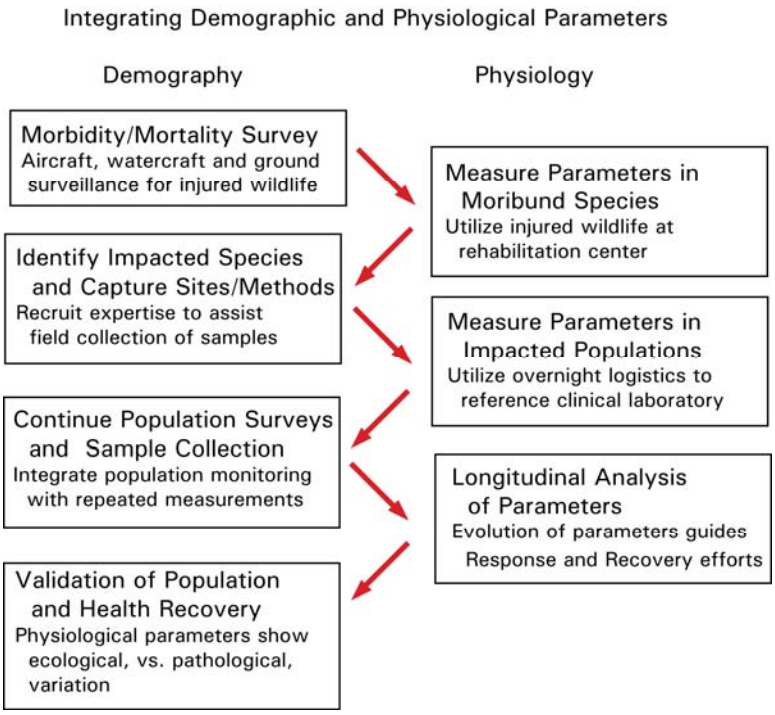


Figure 11. Flow-chart illustrating how integration of physiological parameters into demographic analyses can inform the oil-spill response and restoration process.

The initial morbidity/mortality survey that is part of oil spill search and collection efforts provides a population of animals in rehabilitation that can be assessed utilizing the hematological parameters discussed in this report. The results of these initial tests can then be used to target particular populations that exhibit evidence of physiological abnormalities for further testing in the field. Our results show that measurement of hematocrit, a test that requires small volumes of blood and which is routinely evaluated in the rehabilitation center, may be an effective parameter to identify and target impacted populations. Collection of field samples and submission of these samples to a veterinary clinical laboratory allows for real-time analysis of the effects of oil on physiological systems. Once this data is in hand, it can be correlated with short-term population assessments to refine further sampling techniques. Eventually, this integration of data leads to targeted response and recovery efforts and validation of population and health recovery.

A crucial aspect of this integration process is partnering with a veterinary clinical laboratory that can provide analytical services in "real time" during response and restoration operations. Originally, we anticipated that the national network of commercial and university veterinary clinical services laboratories would be utilized for analyzing the physiological parameters for resource damage assessment. However, two important issues arose concerning the utilization of this network of veterinary clinical laboratories: First, many commercial laboratories give highest priority to veterinary practitioners and some outright refuse to analyze samples for research purposes. We did not explore how these laboratories would respond to requests for analyses related to an oil spill incident. Second, clinical "service" laboratories at university veterinary

schools are amenable to analyzing research samples on a per fee basis. While QC\QA is effective at these laboratories, confidentiality and chain of custody procedures appear to be problematic.

After considerable exploration, we have identified criteria for commercial veterinary services laboratories that will be a crucial for the technology transfer aspects of this research project.
1. Laboratories should have extremely rigorous QC\QA programs. These laboratories typically specialize in preclinical research trials of veterinary products.
2. Laboratories that offer customizable plasma chemistry analyses from a wide array (up to 40) of available analytes provide the flexibility to incorporate parameters that are appropriate or the impacted species and pathogenesis from spilled petroleum. The panel of analytes in our research project was restricted by the small plasma sample volumes available from common terns.
3. Laboratories should be staffed with board-certified avian pathologists that supervise the analyses and are available for consultation for interpretation of results.
4. Laboratories that have: (1) facilities in proximity to at-risk coastal regions; (2) routinely receive specimens via overnight express service; (3) offer expedited services; and (4) use internet communication would be most effective in an oil spill incident.
5. Laboratories with specialized research service teams associated with drug development and Good Laboratory Practices are also trained in chain-of custody procedures. This addresses an important legal aspect of oil spill response activities that may be problematic for typical commercial and university laboratories.
6. These research service teams are familiar with confidentiality and disclosure requirements associated with commercial drug development. This aspect also dovetails nicely with the legal requirements associated with an oil spill response.

We encourage agencies that are responsible for wildlife injury mitigation to identify and contact commercial veterinary clinical laboratories that meet these criteria. Through liaison, they should establish Standard Operating Procedures for collection, submission, analysis and reporting of hematological parameters in order to inform decision makers in a timely fashion..

7.0 Achievement and Dissemination

Presentations made at Inter- and National Conferences:

Natural and oil-spill associated variation in hematocrit in Buzzards Bay, Massachusetts Wildlife Disease Association and the Association of Wildlife Veterinarians, Storrs, CT. 6-10 August 2006

Hematocrit discriminates between natural and oil-spill associated physiological responses Fourth North American Ornithological Conference, Vera Cruz, Mexico. 3-7 October 2006

Manuscripts in preparation

Variability and linkage of demographic parameters of breeding tern populations in the Northeast United States.

Distinguishing between natural and oil-associated demographic responses in seabirds using hematocrit.

References

Alonso-Alvarez, C., M. Ferrer, and A. Velando. 2002. The plasmatic index of body condition in yellow-legged gulls *Larus cachinnans*: a food-controlled experiment. Ibis 144:147-149.

Apanius, V. 1998. Stress and immune defense. Pp. 133-153 in P. J. B. Slater, J. S. Rosenblatt, C. T. Snowden and M. Milinski, eds. Advances in the Study of Behavior. Academic Press, San Diego.

Apanius, V., and I. Nisbet. 2006. Serum immunoglobulin G levels are positively related to reproductive performance in a long-lived seabird, the common tern (*Sterna hirundo*). Oecologia 147:12-23

Austin, O.L. 1944. The status of Tern Island and the Cape Cod terns in 1943. Bird Banding 15:10-27.

Brocklebank, J.C. and D.A. Dickey. 2003. SAS for Forecasting Time Series. SAS Institute, Inc., Cary, NC.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Hillsdale, NJ.

Dawson, R. D., and G. R. Bortolotti. 1997. Are avian hematocrits indicative of condition? American kestrels as a model. Journal of Wildlife Management 61:1297-1306.

Fudge, A.M (*ed*). 2000. Laboratory Medicine Avian and Exotic Pets. Saunders, Philadelphia, PA.

Ghebremeskel, K., T. D. Williams, G. Williams, D. A. Gardners, and M. A. Crawford. 1992. Dynamics of plasma nutrients and metabolites in moulting macaroni (*Eudyptes chrysolophus*) and gentoo (*Pygoscelis papua*) penguins. Comparative Biochemistry and Physiology 101A:301-307.

Golet, G. H., P. E. Seiser, A. D. McGuire, D. D. Roby, J. B. Fischer, K. J. Kuletz, D. B. Irons, T. A. Dean, S. C. Jewett, and S. H. Newman. 2002. Long-term direct and indirect effects of the 'Exxon Valdez' oil spill on pigeon guillemots in Prince William Sound, Alaska. Marine Ecology Progress Series 241:287-304

Hollmen, T., J. C. Franson, M. Hario, S. Sankari, M. Kilpi, and K. Lindstrom. 2001. Use of serum biochemistry to evaluate nutritional status and health of incubating common Eiders (*Somateria mollissima*) in Finland. Physiological and Biochemical Zoology 74:333-342.

Jeffrey, D. A., D. B. Peakall, D. S. Miller, and G. R. Herzberg. 1985. Blood chemistry changes in food-deprived herring gulls. Comparative Biochemistry and Physiology A 81:911-913.

Jessup, D., and F. Leighton. 1996. Oil pollution and petroleum toxicity to wildlife. *In* Noninfectious Diseases of Wildlife (A. Fairbrother, L. Locke and G. Hoff, *eds*). Iowa State University Press, Ames, IA. Pp. 141-157.

Kaneko, J. 1997. Serum proteins and the dysproteinemias. *In* Clinical Biochemistry of Domestic Animals (J. Kaneko, J. Harvey and M. Bruss, eds.). Academic Press, New York, NY. Pp. 117-138

Leighton, F.A. 1993. The toxicity of petroleum oils to birds. Environmental Reviews 1:92-103.

Leighton, F.A., Peakall, D.B., and Butler, R.G. 1983. Heinz body hemolytic anemia from ingestion of crude oil: A primary toxic effect in marine birds. Science 220: 871-873.

Miller, D.S., D.B. Peakall, and W.B. Kinter. 1978. Ingestion of crude oil: sublethal effects in herring gull chicks. Science 199:315-317.

Newman, S., J. Mazet, M. Ziccardi, C. Lieske, D. Fauquier, I. Gardner, J. Zinkl, and M. Christopher. 1999. Haematological changes and anaemia associated with captivity and petroleum exposure in seabirds. Comparative Clinical Pathology 9:60-67.

Newman, S. H., Anderson,D.W., M. H. Ziccardia, J. G. Trupkiewiczc, F. Tseng, M. Christophere, and J. Zinkle. 2000. An experimental soft-release of oil-spill rehabilitated American coots (*Fulica americana*): II. Effects on health and blood parameters. Environmental Pollution 107:293-304.

Nisbet, I.C.T. and J.A. Spendelow. 1999. Contribution of research to management and recovery of the Roseate Tern: review of a twelve-year project. Waterbirds 22: 239-252.

Nisbet, I.C.T. and M. Welton. 1984. Seasonal variations in breeding success of Common Terns: consequences of predation. Colonial Waterbirds 86:53-60.

Peterson, C., S. Rice, J. Short, D. Esler, J. Bodkin, B. Ballachey, and D. B. Irons. 2003. Long-term ecosystem response to the Exxon Valdez Oil Spill. Science 302:2082-2086.

Seiser, P. E., L. K. Duffy, A. D. McGuire, D. Roby, G. H. Golet, and M. A. Litzow. 2000. Comparison of pigeon guillemot, *Cepphus columba*, blood parameters from oiled and unoiled areas of Alaska eight years after the Exxon Valdez oil spill. Marine Pollution Bulletin 40:152-164.

Solberg, W.E. 2001. Establisment and use of reference values. *In* Tietz Fundamentals of Clinical Chemistry (C.A. Burtis and E.R. Ashwood, *eds.*). Saunders, Philadelphia, PA. Pp. 251-261.

Thompson, W.L., G. White, and C. Gowan. 1998. Monitoring Vertebrate Populations. Academic Press, San Diego.

Tims, J., I.C.T. Nisbet, M.S. Friar,. C. Mostello, and J.J. Hatch. 2004. Comparison of characteristics and performance of Common Terns at old and newly-established breeding colonies. Waterbirds 37: 312-319.

Westfall, P.H. and S.S. Young. 1993. Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment. Wiley and Sons, New York, NY.

Westfall, P.H., R.D. Tobias, R. Dror, R.D. Wolfinger, and Y. Hochberg. 1999.Multiple Comparisons and Multiple Tests Using the SAS System. SAS Institute Inc., Cary, NC.

Yamato, O., I. Goto, and Y Maede. 1996. Hemolytic anemia in wild seaducks caused by caused by marine oil pollution. Journal of Wildlife Diseases 32:381-384.

Young, D.S. and E.W. Bermes. 2001. Specimen collection and other preanalytical variables. *In* Tietz Fundamentals of Clinical Chemistry (C.A. Burtis and E.R. Ashwood, *eds.*). Saunders, Philadelphia, PA. Pp. 30-54.

Zar, J.H. 1996. Biostatistical Analysis. Prentice Hall, Upper Saddle River, NJ.

Appendix Tables and Figures

Table A1. Best fit statistical models for breeding population size. The data were ln-transformed in order to stabilize the residual (error) variance. Years refers to the time period analyzed in each model. The intercept parameter is the mean value of the first year of the time period. The year parameter is indexed to the year within the time period, e.g. for years 1971-1980, 1971 is YEAR=0, 1972 is Year=1, 1973 is Year=2, etc. The Year^2 parameter is a quadratic (year x year) form of the year parameter. The ρ parameter is the first-order autocorrelation coefficient. Parameter estimates and standard errors are in the ln-transformed scale; to backtransform x, use exp^x. t is the student t test statistic for the test that the parameter estimate is not zero. DW is the Durbin-Watson test statistic for the significance of the autocorrelation coefficient. P is the significance level of the t- or DW test statistics. Model df is the regression model degrees of freedom. Model F is the F-ratio test statistic for the regression model. Model P is the significance level of the F-ratio test statistic. Model r^2 is the coefficient of determination. Model RMSE is the root mean square error of the regression model. Model CV is the coefficient of variation expressed as a percentage. Regression models were based on ordinary least squares methodology, except where noted by AR in the Years column, which was a re-analysis of the same time period using a regression model with autoregressive errors. NA = Not Applicable.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Common Tern Monomoy Island	1971-1980	Intercept	8.311	0.110	75.36	<0.0001	1,8	4.18	0.075	0.343	0.188	2.21
		Year	0.042	0.021	2.04	0.075						
		ρ	0.051	0.377	1.55	0.11						
	1980-1988	Intercept	10.960	1.103	9.94	<0.0001	1,6	9.30	0.024	0.608	0.563	7.35
		Year	-0.265	0.087	-3.05	0.022						
		ρ	-0.123	0.444	1.81	0.21						

Table A1, cont.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Common Tern Bird Island	1970-1989	Intercept	6.193	0.047	132.55	<0.0001	1,20	469.14	<0.0001	0.959	0.131	1.87
		Year	0.095	0.004	21.66	<0.0001						
		ρ	0.447	0.205	0.96	0.0015						
	AR	Intercept	6.182	0.070	88.31	<0.0001	NA	NA	NA	0.969	0.118	1.68
		Year	0.098	0.006	15.18	<0.0001						
	1989-2002	Intercept	8.200	0.021	383.12	<0.0001	NA	NA	NA	NA	0.080	0.98
		ρ	0.048	0.288	1.84	0.38						
Roseate Tern Bird Island	1970-1986	Intercept	7.498	0.091	82.40	<0.0001	1,15	17.20	0.0009	0.534	0.196	2.51
		Year	0.040	0.010	4.15	0.0009						
		ρ	-0.036	0.267	1.64	0.14						
	1986-2002	Intercept	8.186	0.081	100.88	<0.0001	1,15	28.03	<0.0001	0.651	0.175	2.23
		Year	-0.046	0.009	-5.29	<0.0001						
		ρ	0.029	0.267	1.42	0.060						

Table A1, cont.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Roseate Tern Flakner Island	1978-1993	Intercept	5.675	0.052	108.17	<0.0001	NA	NA	NA	NA	0.210	3.70
		ρ	0.249	0.251	2.41	0.20						
	1993-2002	Intercept	5.701	0.061	93.55	<0.0001	1,7	68.85	<0.0001	0.908	0.126	2.36
		$Year^2$	-0.016	0.002	-8.30	<0.0001						
		ρ	0.168	0.402	1.38	0.067						
	AR	Intercept	5.702	0.075	76.34	<0.0001	NA	NA	NA	0.911	0.133	2.49
		$Year^2$	-0.017	0.002	-7.11	0.0004						

Table A2. Correlation of (ln-transformed) breeding population size among populations. In each cell, the top number is the Pearson correlation coefficient (r), the middle number is the signficance (P-value), and the bottom number is the sample size (N). Values above the diagonal are original data, values below the diagonal are the residuals from the regression models.

	Common Tern Monomoy Island	Common Tern Bird Island	Roseate Tern Bird Island	Roseate Tern Falkner Island
Common Tern Monomoy Island	-------	-0.729 0.0009 17	-0.418 0.095 17	-0.336 0.34 10
Common Tern Bird Island	0.119 0.64 17	-------	0.183 0.30 33	-0.293 0.15 25
Roseate Tern Bird Island	0.267 0.29 17	0.293 0.098 33	-------	0.654 0.0004 25
Roseate Tern Falkner Island	-0.518 0.12 10	0.175 0.40 25	0.278 0.17 25	-------

Table A3. Best fit statistical models for median egg-laying date. Format follows Table A1, except data were not ln-transformed.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Common Tern Monomoy Island	1971-1980	Intercept	36.14	2.41	14.96	<0.0001	1,7	15.99	0.0052	0.696	3.33	12.1
		Year	-1.74	0.43	-4	0.0052						
		ρ	0.278	0.392	1.34	0.055						
	1980-1986	Intercept	24.39	3.21	7.60	0.0006	1,5	9.06	0.030	0.645	4.71	14.5
		Year	2.68	0.89	3.01	0.03						
		ρ	0.467	0.442	2.91	0.19						
Bird Island	1970-1980	Intercept	24.55	1.1	22.31	<0.0001	1,9	32.71	0.0003	0.784	1.95	10.1
		Year	-1.06	0.19	-5.72	0.0003						
		ρ	-0.05	0.387	1.58	0.13						
	1980-1992	Intercept	11.08	1.51	7.32	<0.0001	1,11	31.08	0.0002	0.739	2.88	15.8
		Year	1.19	0.21	5.57	0.0002						
		ρ	-0.454	0.327	1.02	0.009						
Common Tern		Intercept	10.60	1.88	5.64	0.0005	NA	NA	NA	0.852	2.55	
		Year	1.24	0.271	4.58	0.0018						
	1993-2002	Intercept	19.50	0.86	22.69	<0.00001	NA	NA	NA	NA	2.72	13.9
		ρ	-0.145	0.401	1.38	0.15						

Table A3, cont.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Roseate Tern Bird Isalnd	1980-1993	Intercept	19.4	1.18	16.46	<0.0001	1,9	38.44	0.0002	0.81	1.73	6.7
		Year	0.86	0.14	6.20	0.0002						
		ρ	-0.189	0.347	2.36	0.4						
	1993-2002	Intercept	30.07	1.92	15.63	<0.0001	1,8	4.88	0.058	0.379	2.82	10.7
		Year	-0.69	0.31	-2.21	0.058						
Roseate Tern Falkner Island	1978-1992	Intercept	21.14	1.47	14.41	<0.0001	1,13	36.34	<0.0001	0.737	2.98	10.4
		Year	1.08	0.18	6.03	<0.0001						
		ρ	0.278	0.277	1.21	0.024						
		Intercept	21.52	1.89	11.39	<0.0001	NA	NA	NA	0.762	2.95	
		Year	1.05	0.226	4.61	0.0006						
	1992-2002	Intercept	34.86	2.01	17.37	<0.0001	1,9	8.22	0.019	0.477	3.56	11.96
		Year	-0.97	0.34	-2.87	0.019						
		ρ	-0.254	0.342	2.29	0.45						

Table A4. Correlations of median egg-laying date among populations. Format as in Table A2.

	Common Tern Monomoy Island	Common Tern Bird Island	Roseate Tern Bird Island	Roseate Tern Falkner Island
Common Tern Monomoy Island	-------	0.079 0.77 15	0.791 0.20 4	0.281 0.46 9
Common Tern Bird Island	-0.305 0.26 15	-------	0.589 0.0062 20	0.679 0.0002 25
Roseate Tern Bird Island	-0.396 0.60 4	0.165 0.48 20	-------	0.613 0.0041 20
Roseate Tern Falkner Island	0.104 0.79 9	0.213 0.30 25	0.184 0.43 20	-------

Table A5. Best fit statistical models for mean clutch size. Format follows Table A1, except data were not ln-transformed. RT refers to Roseater Tern, BI refers to Bird Island, and FI refers to Falkner Island.

Popula-tion	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Common Tern Monomoy Islnad	1972-1981	Intercept	2.554	0.070	36.36	<0.0001	1,8	6.68	0.032	0.455	0.120	4.41
		Year	0.034	0.013	2.58	0.0324						
		ρ	0.074	0.377	1.57	0.12						
	1983-1987	Intercept	2.888	0.078	36.77	<0.0001	1,3	13.54	0.035	0.818	0.101	3.82
		Year	-0.118	0.032	-3.68	0.035						
Common Tern Bird Island	1970-1985	Intercept	2.905	0.012	241.25	<0.0001	NA	NA	NA	NA	0.048	1.66
		ρ	0.126	0.265	2.13	0.4						
	1985-1990	Intercept	2.884	0.042	68.56	<0.0001	1,4	83.38	0.0008	0.954	0.058	2.26
		Year	-0.127	0.014	-9.13	0.0008						
		ρ	-0.227	0.562	2.4	0.47						
	1990-2002	Intercept	2.389	0.054	44.51	<0.0001	NA	NA	NA	NA	0.194	8.10
		ρ	-0.502	0.261	2.73	0.081						
RT BI	1980-2002	Intercept	1.828	0.016	114.41	<0.0001	NA	NA	NA	NA	0.075	4.10
		ρ	-0.071	0.223	2.13	0.38						
RT FI	1980-2002	Intercept	1.828	0.018	103.76	<0.0001	NA	NA	NA	NA	0.088	4.82
		ρ	0.254	0.202	2.51	0.096						

Table A6. Correlations of mean clutch size among populations. Format as in Table A2.

	Common Tern Monomoy Island	Common Tern Bird Island	Roseate Tern Bird Island	Roseate Tern Falkner Island
Common Tern Monomoy Island	-------	0.617 0.014 15	-0.383 0.45 6	-0.031 0.93 9
Common Tern Bird Island	0.083 0.76 15	-------	0.258 0.25 22	0.055 0.79 25
Roseate Tern Bird Island	-0.754 0.083 6	0.231 0.30 22	-------	0.358 0.10 22
Roseate Tern Falkner Island	-0.530 0.14 9	0.375 0.065 25	0.358 0.10 22	-------

Table A7. Best fit statistical models for mean A-egg mass. Format follows Table A1, except data were not ln-transformed.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Common Tern Monomoy Island	1972-1981	Intercept	20.54	0.15	134.85	<0.0001	1,6	9.47	0.028	0.654	0.22	1.07
		Year	0.13	0.04	3.08	0.028						
		ρ	0.609	0.396	3.18	0.084						
	1983-1986	Intercept	21.10	0.048	439.62	<0.0001	1,2	153.68	0.0064	0.987	0.06	0.28
		Year	-0.32	0.026	-12.4	0.0064						
		ρ	-0.264	0.965	2.04	0.11						
Common Tern Bird Island	1981-2002	Intercept	21.53	0.16	135.71	<0.0001	1,20	9.21	0.0066	0.315	0.21	1.00
		Year	-0.021	0.007	-3.03	0.0066						
		ρ	0.277	0.22	1.36	0.033						
	AR	Intercept	21.49	0.21	104.84	<0.0001	NA	NA	NA	0.374	0.21	0.99
		Year	-0.02	0.009	-2.16	0.044						
Roseater Tern Bird Island	1980-1990	Intercept	20.57	0.05	455.05	<0.0001	NA	NA	NA	NA	0.14	0.66
		ρ	-0.113	0.376	1.47	0.2						
	1987-2002	Intercept	21.01	0.08	250.53	<0.0001	1,4	44.56	0.0026	0.918	0.09	0.44
		Year	-0.14	0.02	-6.68	0.0026						
		ρ	0.015	0.577	1.66	0.12						

Table A7, cont.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Roseate Tern Falkner Island	1988-2002	Intercept	20.69	0.06	357.38	<0.0001	NA	NA	NA	NA	0.22	0.08
		ρ	0.025	0.277	1.85	0.38						

Table A8. Correlations of mean A-egg mass among populations. Format as in Table A2.

	Common Tern Monomoy Island	Common Tern Bird Island	Roseate Tern Bird Island	Roseate Tern Falkner Island
Common Tern Monomoy Island	-------	-0.703 0.078 7	NA	NA
Common Tern Bird Island	0.017 0.98 4	-------	0.355 0.21 14	-0.158 0.57 15
Roseate Tern Bird Island	NA	0.548 0.042 14	-------	-0.375 0.32 9
Roseate Tern Falkner Island	NA	-0.229 0.41 15	-0.344 0.36 9	-------

Table A9. Best fit statistical models for mean clutch mass. Format follows Table A1, except data were not ln-transformed. CT-MI refers to Common Terns on Monomoy Island, CT-BI refers to Common Terns on Bird Island, RT-FI refers to Roseate Terns on Falkner Island.

Popula-tion	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
CT MI	1972-1981	Intercept	60.35	0.32	186.39	<0.0001	1,6	28.34	0.0018	0.825	0.50	0.81
		Year	0.41	0.08	5.32	0.0018						
		ρ	0.161	0.441	1.48	0.088						
CT BI	1972-2002	Intercept	63.81	0.39	163.37	<0.0001	1,23	22.51	<0.0001	0.495	1.05	1.69
		Year	-0.11	0.02	-4.74	<0.0001						
		ρ	0.201	0.209	1.97	0.38						
Roseate Tern Bird Island	1980-1990	Intercept	39.91	0.13	295.81	<0.0001	NA	NA	NA	NA	0.40	1.01
		ρ	0.002	0.378	1.43	0.18						
	1997-2002	Intercept	40.67	0.22	180.86	<0.0001	1,4	13.30	0.022	0.769	0.24	0.60
		Year	-0.21	0.06	-3.65	0.022						
		ρ	-0.152	0.571	2.09	0.3						
RT FI	1988-2002	Intercept	40.25	0.15	262.81	<0.0001	NA	NA	NA	NA	0.59	1.47
		ρ	0.144	0.275	2.10	0.42						

Table A10. Correlations of mean clutch mass among populations. Format as in Table A2

	Common Tern Monomoy Island	Common Tern Bird Island	Roseate Tern Bird Island	Roseate Tern Falkner Island
Common Tern Monomoy Island	-------	-0.096 0.79 10	NA	NA
Common Tern Bird Island	0.542 0.21 7	-------	0.123 0.69 13	0.172 0.65 9
Roseate Tern Bird Island	NA	0.192 0.53 13	-------	-0.316 0.40 9
Roseate Tern Falkner Island	NA	0.145 0.71 9	-0.173 0.65 9	-------

Table A11. Best fit statistical models for mean productivity. Format follows Table A1, except data were not ln-transformed.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Common Tern Monomoy Is.land	1972-1982	Intercept	1.729	0.250	6.91	<0.0001	1,9	12.58	0.0063	0.583	0.568	52.4
		$YEAR^2$	-0.019	0.005	-3.55	0.0063						
		ρ	0.081	0.332	1.47	0.088						
	1982-1987	Intercept	0.03	0.021	1.45	0.20	NA	NA	NA	NA	0.051	168.7
Common Tern Bird Island	1970-1983	Intercept	1.783	0.120	14.90	<0.0001	1,12	7.42	0.018	0.382	0.236	11.5
		Year	0.043	0.016	2.72	0.018						
		ρ	-0.437	0.271	2.82	0.091						
	1983-1991	Intercept	2.414	0.064	37.94	<0.0001	1,7	247.36	<0.0001	0.973	0.104	6.6
		Year	-0.21	0.013	-15.73	<0.0001						
		ρ	0.223	0.398	1.37	0.062						
	1991-2002	Intercept	0.969	0.100	9.65	<0.0001	NA	NA	NA	NA	0.348	35.9
		ρ	0.089	0.315	1.33	0.11						
Roseate Tern Bird Island	1986-2002	Intercept	1.277	0.037	34.57	<0.0001	NA	NA	NA	NA	0.152	11.9
		ρ	-0.047	0.258	1.77	0.31						

Table A11, cont.

Population	Years	Parameter					Model					
		Name	Est.	SE	t/DW	P	df	F	P	r^2	RMSE	CV
Roseate Tern Falkne Island	1978-2002	Intercept	1.176	0.059	20.01	<0.0001	1,23	28.04	<0.0001	0.549	0.198	20.9
		$YEAR^2$	-0.001	0.0002	-5.29	<0.0001						
		ρ	0.27	0.205	1.30	0.019						
	AR	Intercept	1.176	0.077	15.32	<0.0001	NA	NA	NA	0.588	0.194	20.5
		$YEAR^2$	-0.001	0.0003	-4.28	0.0003						

Table A12. Correlations of mean productivity among populations. Format as in Table A2.

	Common Tern Monomoy Island	Common Tern Bird Island	Roseate Tern Bird Island	Roseate Tern Falkner Island
Common Tern Monomoy Island	-------	-0.107 0.69 16	NA	-0.503 0.14 10
Common Tern Bird Island	0.225 0.40 16	-------	0.702 0.0012 18	0.781 <0.0001 25
Roseate Tern Bird Island	NA	0.737 0.0007 17	-------	0.547 0.019 18
Roseate Tern Falkner Island	0.546 0.10 10	0.478 0.016 25	0.568 0.018 17	-------

Table A13. Correlations among demographic parameters of Common Terns on Monomoy Island. In each cell, the top number is the Pearson correlation coefficient (r), the middle number is the signficance (P-value), and the bottom number is the sample size (N). Values above the diagonal are original data, values below the diagonal are the residuals from the regression models.

	POPULATION	LAYDATE	CLUTCHSIZE	EGGMASS	CLUTCHMASS	PRODUCTIVITY
POPULATION (breeding population size)	--------	-0.764 0.0009 15	0.309 0.26 15	0.673 0.023 11	0.304 0.36 11	0.503 0.047 16
LAYDATE (median egg-laying date)	-0.296 0.28 15	------	-0.498 0.069 14	-0.579 0.062 11	-0.491 0.12 11	-0.384 0.16 15
CLUTCHSIZE (mean clutch size)	-0.179 0.52 15	-0.082 0.78 14	------	0.680 0.021 11	0.690 0.019 11	0.076 0.78 15
EGGMASS (mean A-egg mass)	-0.096 0.78 11	0.286 0.39 11	0.007 0.98 11	------	0.554 0.097 10	0.290 0.38 11
CLUTCHMASS (mean clutch mass)	0.523 0.18 8	0.129 0.76 8	-0.026 0.95 8	0.301 0.51 7	-------	-0.475 0.14 11
PRODUCTIVITY	-0.025 0.93 16	-0.311 0.26 15	0.292 0.29 15	0.460 0.16 11	0.025 0.95 8	-------

Table A14. Correlations among demographic parameters of Common Terns on Bird Island. Format as in Table A13.

	POPULATION	LAYDATE	CLUTCHSIZE	EGGMASS	CLUTCHMASS	PRODUCTIVITY
POPULATION	-------	0.078	-0.804	-0.373	-0.654	-0.751
		0.66	<0.0001	0.055	0.0004	<0.0001
		33	33	27	25	33
LAYDATE	-0.399	-------	-0.401	-0.198	-0.172	-0.482
	0.021		0.021	0.32	0.4108	0.0045
	33		33	27	25	33
CLUTCHSIZE	0.153	-0.131	------	0.338	0.561	0.917
	0.39	0.46		0.084	0.0035	<0.0001
	33	33		27	25	33
EGGMASS	0.273	-0.582	-0.132	-------	0.593	0.363
	0.21	0.0044	0.55		0.0046	0.063
	22	22	22		21	27
CLUTCHMASS	-0.111	-0.059	-0.303	0.501	-------	0.553
	0.59	0.77	0.14	0.048		0.0041
	25	25	25	16		25
PRODUCTIVITY	0.018	-0.233	0.623	-0.095	-0.130	-------
	0.92	0.19	0.0001	0.67	0.53	
	33	33	33	22	25	

Table A15. Correlations among demographic parameters of Roseate Terns on Bird Island. Format as in Table A13.

	POPULATION	LAYDATE	CLUTCHSIZE	EGGMASS	CLUTCHMASS	PRODUCTIVITY
POPULATION	-------	-0.068	-0.094	0.397	0.136	0.385
		0.77	0.67	0.14	0.62	0.11
		20	22	15	15	18
LAYDATE	-0.699	-------	-0.011	-0.281	-0.298	-0.290
	0.0006		0.96	0.35	0.32	0.24
	20		19	13	13	18
CLUTCHSIZE	-0.145	0.031	-------	0.084	0.360	0.515
	0.52	0.90		0.77	0.20	0.028
	22	19		14	14	18
EGGMASS	-0.384	0.161	0.154	-------	0.838	0.816
	0.16	0.60	0.59		<0.0001	0.0022
	15	13	14		15	11
CLUTCHMASS	-0.282	0.117	0.444	0.814	-------	0.804
	0.31	0.70	0.11	0.0002		0.0029
	15	13	14	15		11
PRODUCTIVITY	0.425	-0.359	0.528	0.280	0.339	-------
	0.0885	0.16	0.029	0.43	0.34	
	17	17	17	10	10	

Table A16. Correlations among demographic parameters of Roseate Terns on Falkner Island. Format as in Table A13.

	POPULATION	LAYDATE	CLUTCHSIZE	EGGMASS	CLUTCHMASS	PRODUCTIVITY
POPULATION	-------	0.037	0.139	0.257	0.204	0.538
		0.86	0.50	0.35	0.46	0.0055
		25	25	15	15	25
LAYDATE	-0.128	-------	-0.101	0.063	0.034	-0.327
	0.54		0.63	0.82	0.90	0.11
	25		25	15	15	25
CLUTCHSIZE	0.131	-0.409	-------	0.445	0.664	0.324
	0.53	0.042		0.096	0.0069	0.11
	25	25		15	15	25
EGGMASS	0.239	-0.172	0.446	-------	0.819	0.565
	0.39	0.54	0.096		0.0002	0.028
	15	15	15		15	15
CLUTCHMASS	0.158	-0.172	0.664	0.819	-------	0.526
	0.57	0.54	0.0069	0.0002		0.043
	15	15	15	15		15
PRODUCTIVITY	0.147	-0.664	0.509	0.574	0.501	-------
	0.48	0.0003	0.0094	0.025	0.057	
	25	25	25	15	15	

Table A17. Sample sizes (N), means, and standard errors of the mean (SEM) for physiological parameters by year, and some case year by site combinations. B and M refer to Bird and Monomoy Islands, respectively, when following Year. HOV refers to the test for homogeneity of variances: df refers to degrees of freedom, F refers to the F-ratio statistics, and P refers to the significance level. ANOVA refers to the analysis of variance test for significantly different means: df refers to the degrees of freedom (non-integer values represent adjustment for unequal variances), F refers to the F-ratio statistic or Welch's F ratio statistic (bold), P refers to the significance level, and r^2 is the coeeficient of detremination. PVC = hematocrit, TP = total protein, ALB = albumin, UA = uric acid, BUN = blood urea nitrogen, TG = triglycerides, CHL = cholesterol, BA = bile acids, LDH = lactate dehydrogenase, CK = creatine kinase, AST = aspartate aminotransferase. The HOV and ANOVA analyses of total protein, uric acid, triglycerides, lactate dehydrogenase, creatine kinase, aspartate aminotransferase were performed on log-transformed data to stabilize the variance (where possible) and to normalize the distribution of residuals.

					HOV			ANOVA			
Parameter	Year	N	Mean	SEM	df	F	P	df	F	P	r^2
PCV	1999	92	0.506	0.032	4,302	4.41	0.0018	4,134.2	**24.75**	<0.0001	0.234
	2002	58	0.463	0.041							
	2003	75	0.451	0.056							
	2004	42	0.496	0.031							
	2005	40	0.470	0.024							
TP	1999	104	5.51	0.068	4,291	2.96	0.020	4,123.4	**25.56**	<0.0001	0.275
	2002	66	5.07	0.055							
	2003	47	4.90	0.058							
	2004	41	4.75	0.073							
	2005	38	4.69	0.056							

Table A17, cont.

					HOV			ANOVA			
Parameter	Year	N	Mean	SEM	df	F	P	df	F	P	r^2
ALB	1999	88	2.22	0.021	4,257	1.72	0.15	4,257	36.80	<0.0001	0.364
	2002	65	1.92	0.033							
	2003	37	2.15	0.068							
	2004	35	1.89	0.026							
	2005	37	1.70	0.032							
UA	1999	106	11.84	0.42	4,289	0.48	0.75	4,289	4.52	0.0015	0.059
	2002	58	9.41	0.43							
	2003	47	10.35	0.49							
	2004	41	11.31	0.54							
	2005	42	10.41	0.54							
BUN	1999	68	10.70	0.49	5, 245	0.77	0.57	5,245	4.95	0.0002	0.091
	2002	56	8.05	0.86							
	2003	46	8.23	0.46							
	2004	40	8.22	0.42							
	2005B	22	9.63	0.76							
	2005M	20	6.31	0.41							
TG	1999	106	157.7	8.8	5,338	0.50	0.77	5,338	14.81	<0.0001	0.180
	2002	127	104.7	6.5							
	2003	33	121.5	16.3							
	2004	42	109.3	6.8							
	2005	36	102.0	5.9							

Table A17. cont.

					HOV			ANOVA			
Parameter	Year	N	Mean	SEM	df	F	P	df	F	P	r^2
CHL	1999	106	314.1	4.7	5,337	1.62	0.15	5,337	16.70	<0.0001	0.199
	2002	121	308.3	5.1							
	2003	33	306.0	9.7							
	2004	42	237.2	4.8							
	2005B	21	274.4	10.5							
	2005M	20	296.3	8.4							
BA	1999	89	31.20	2.51	3,200	0.69	0.32	3,200	0.08	0.9733	0.001
	2002	6	16.26	2.22							
	2003	33	33.13	4.49							
	2004	42	33.01	5.65							
	2005	40	31.30	3.22							
LDH	1999	81	1446	54	5,234	3.44	0.0051	5,234	77.35	<0.0001	0.613
	2002	39	300	30							
	2003	33	883	96							
	2004	42	576	42							
	2005B	21	843	75							
	2005M	20	560	58							

Table A17. cont.

					HOV			ANOVA			
Parameter	Year	N	Mean	SEM	df	F	P	df	F	P	r^2
CK	1999	104	752.2	74.0	5,259	2.56	0.028	5,77.77	**11.99**	<0.0001	0.216
	2002	46	912.9	124.3							
	2003	33	424.2	61.1							
	2004	42	424.5	41.2							
	2005B	21	970.7	96.4							
	2005M	20	689.9	112.8							
AST	1999	99	183.2	8.86	4,209	0.24	0.91	4,209	14.95	<0.0001	0.223
	2002	6	247.6	53.01							
	2003	33	180.0	24.29							
	2004B	21	86.3	9.00							
	2004M	21	136.3	16.59							
	2005	40	144.2	9.18							

Figure A1. Breeding Population Size for Common Terns on Monomoy Island (a) and Bird Island (b), and Roseate Terns on Bird Island (c) and Falkner Island (d). Dots represent data, thick lines represent predicted values, and thin lines represent 95% prediction intervals from the regression models described in Table A1.

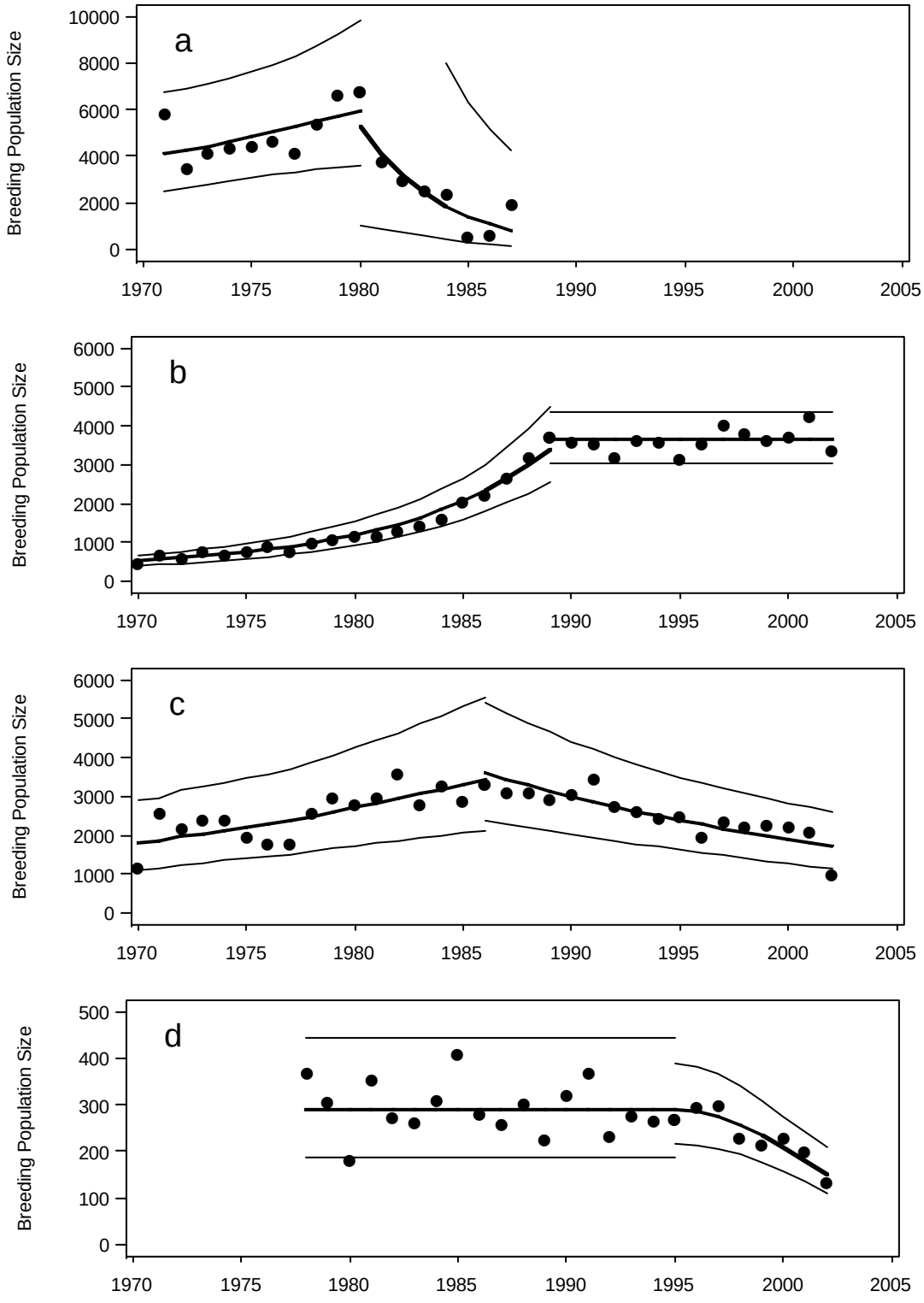


Figure A2. Power curves for the sample required to detect a significant difference ($\alpha = 0.05$) from a given mean, expressed as the proportional difference between means, for CV = 5 (a),10 (b), and 50 % (c). Power refers to 1- β.

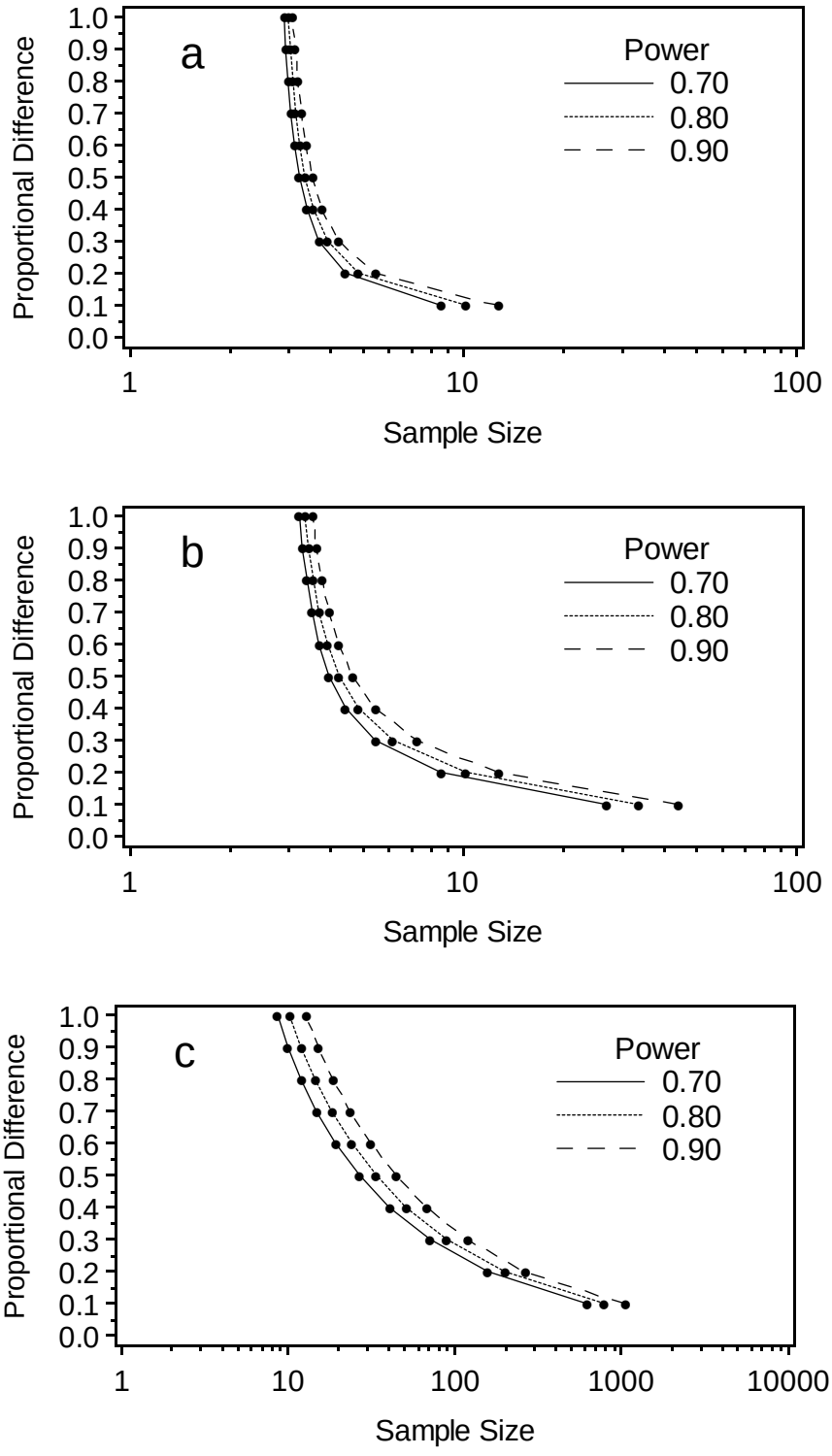


Figure A3. Median egg-laying date for Common Terns on Monomoy Island (a) and Bird Island (b), and Roseate Terns on Bird Island (c) and Falkner Island (d). Date is expressed as the number days after 1 May. Lines and symbols as in Figure A1 based on the regression models shown in Table A3.

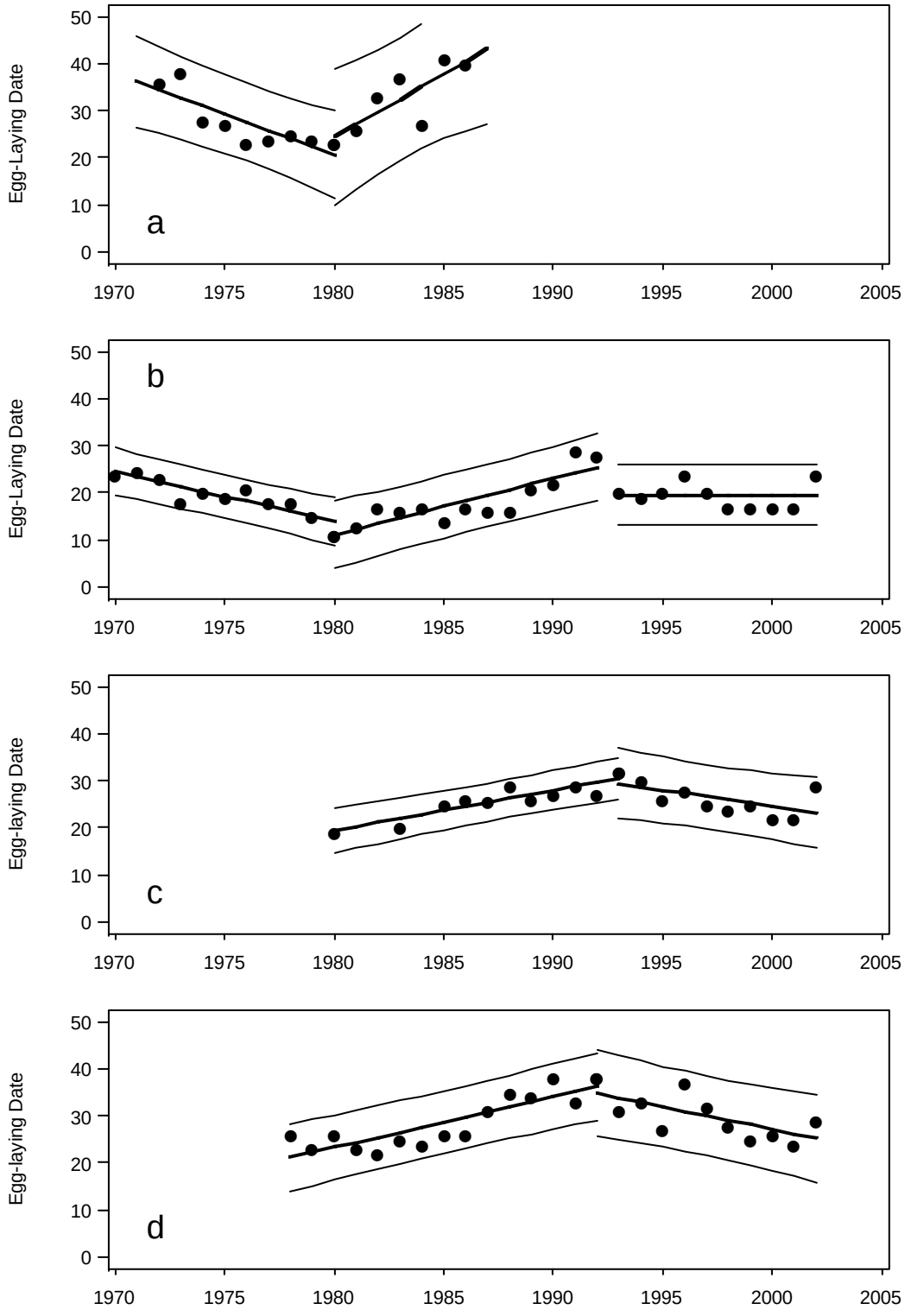


Figure A4. Mean clutch size for Common Terns on Monomoy Island (a) and Bird Island (b), and Roseate Terns on Bird Island (c) and Falkner Island (d). Lines and symbols as in Figure A1 based on the regression models shown in Table A5.

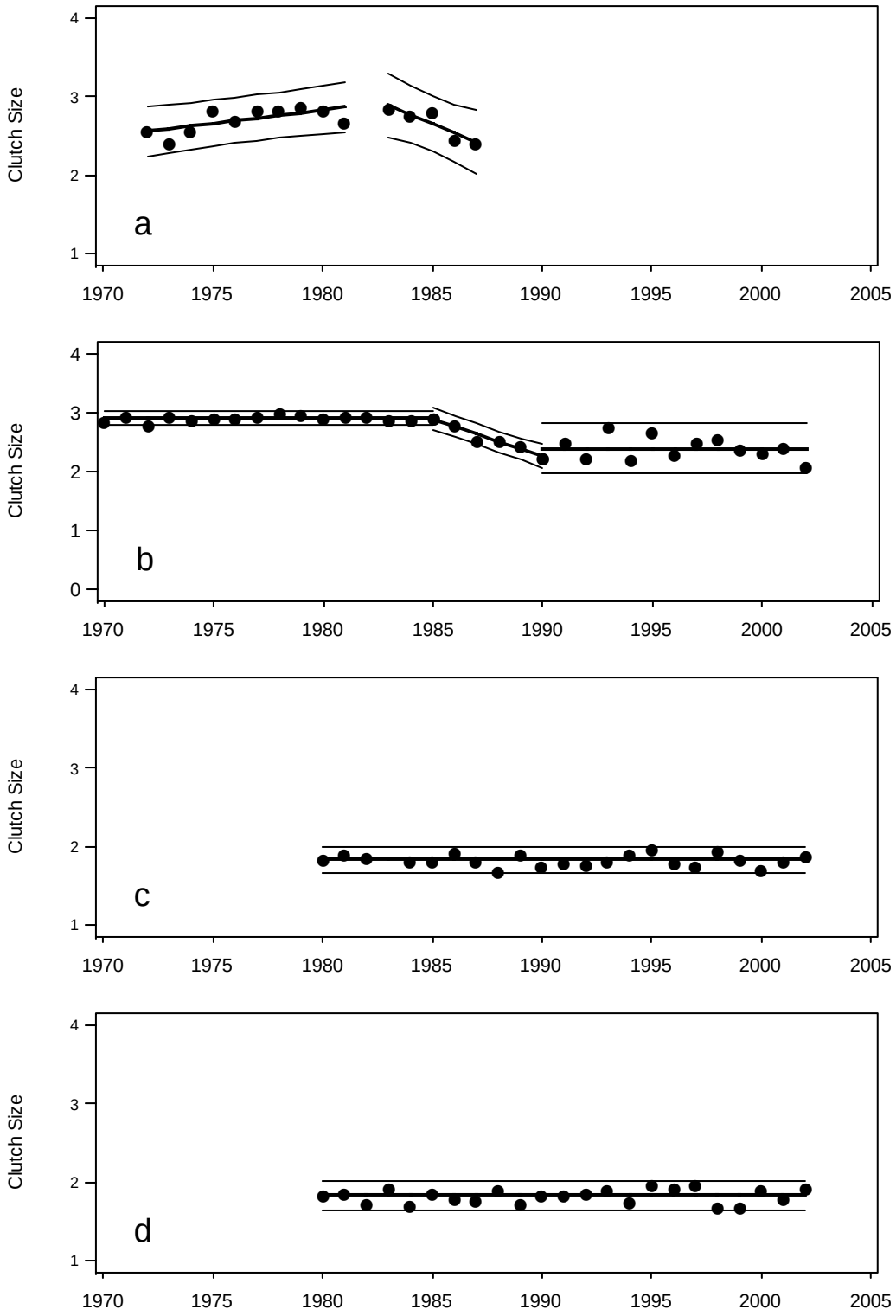


Figure A5. Mean A-egg mass for Common Terns on Monomoy Island (a) and Bird Island (b), and Roseate Terns on Bird Island (c) and Falkner Island (d). Lines and symbols as in Figure A1 based on the regression models shown in Table A7.

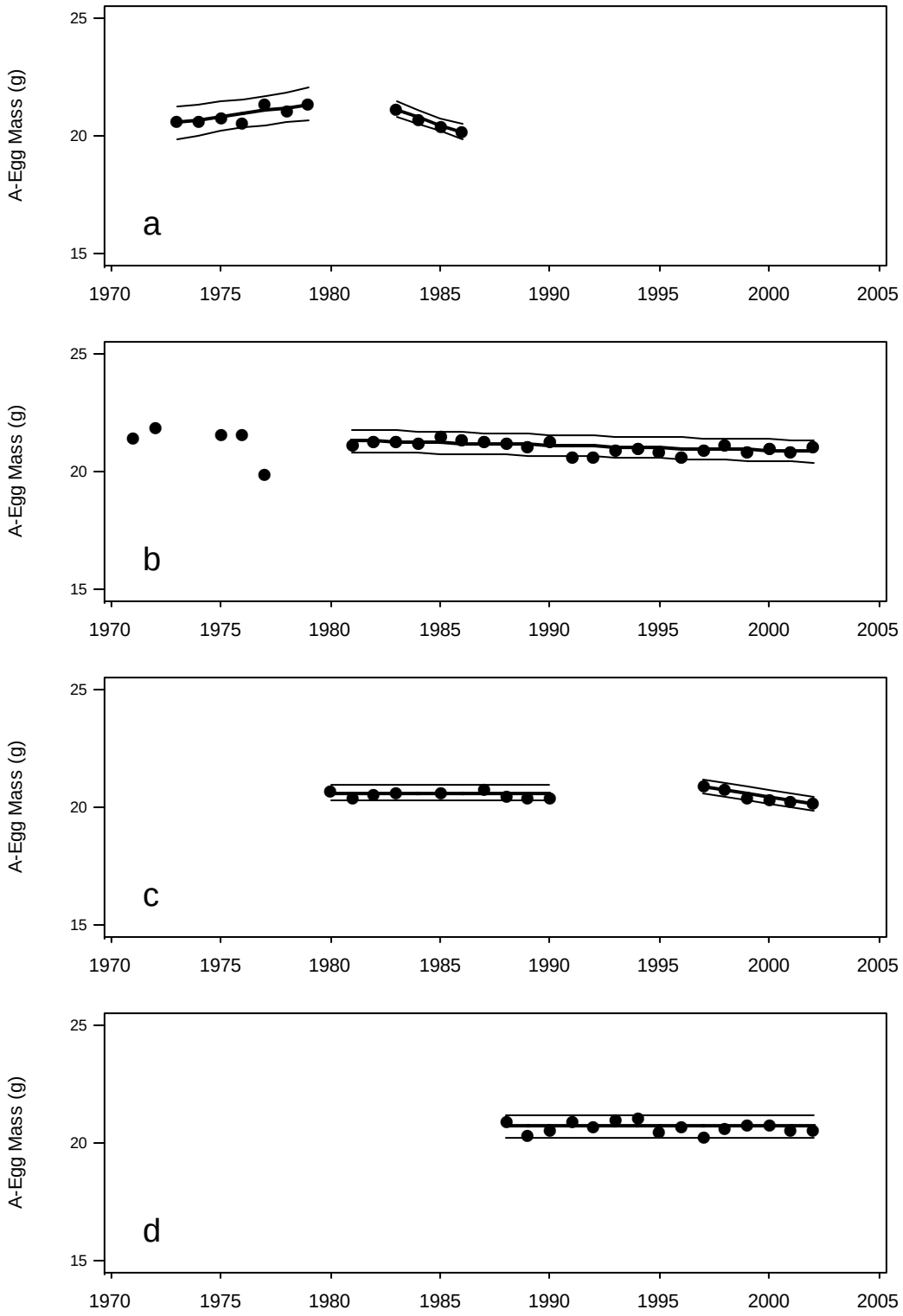


Figure A6. Mean clutch mass for Common Terns on Monomoy Island (a) and Bird Island (b), and Roseate Terns on Bird Island (c) and Falkner Island (d). Lines and symbols as in Figure A1 based on the regression models shown in Table A9.

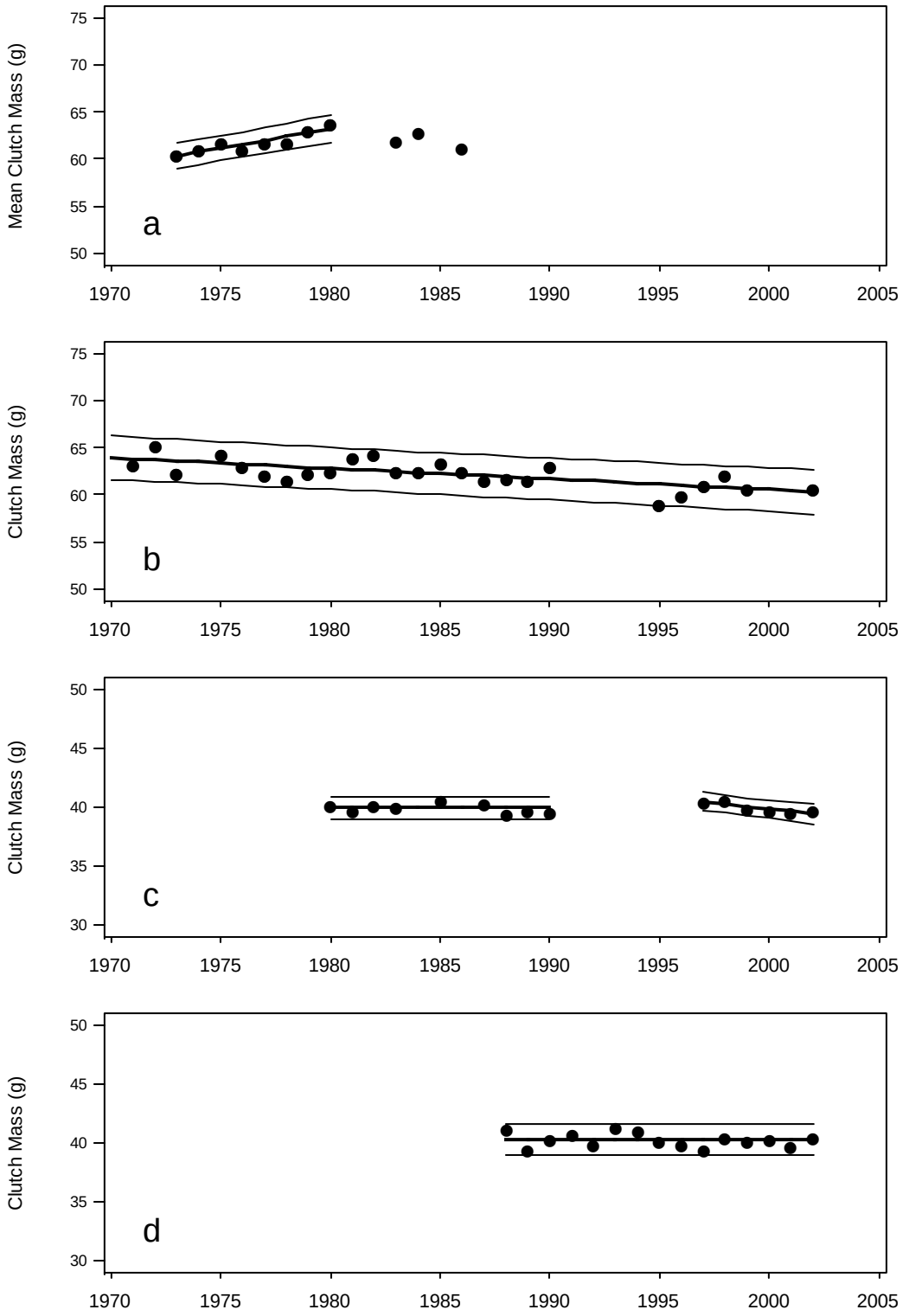


Figure A7. Mean productivity for Common Terns on Monomoy Island (a) and Bird Island (b), and Roseate Terns on Bird Island (c) and Falkner Island (d). Lines and symbols as in Figure A1 based on the regression models shown in Table A11.

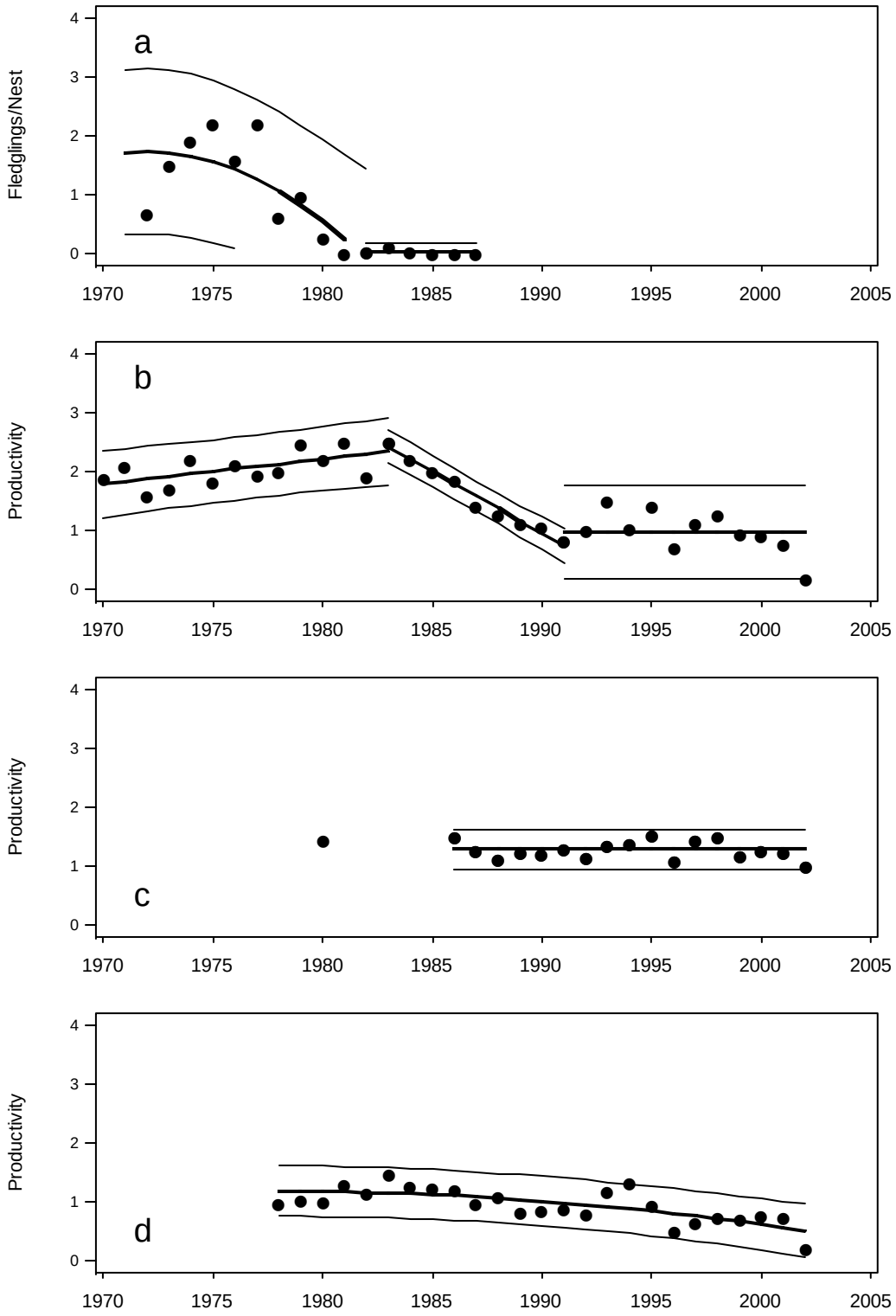
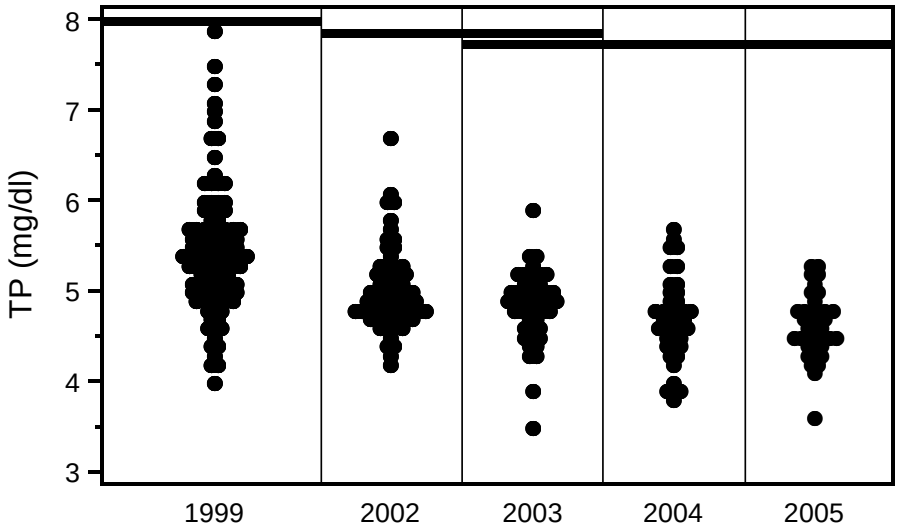


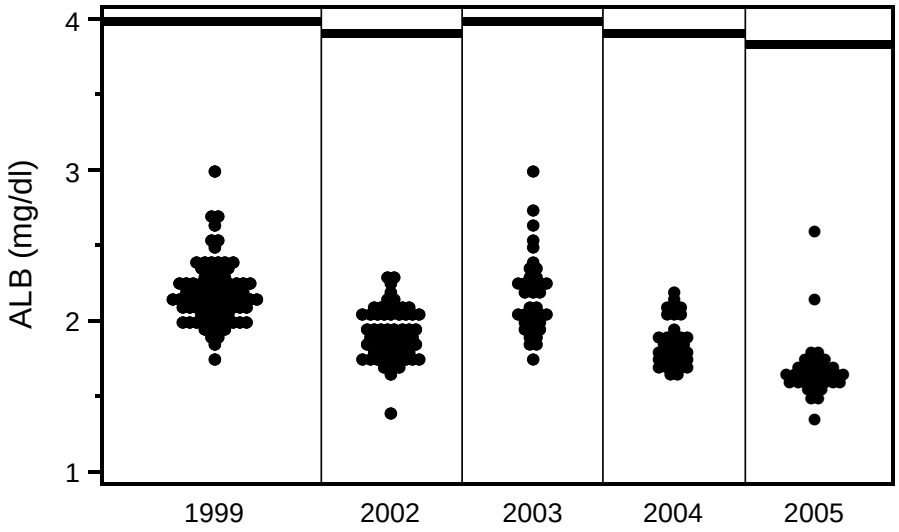
Figure A8. Distribution of total protein (TP) concentrations as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The Biuret reaction using the method of Flack and Woollen (1984; Cat. No. 1700-250, Thermo-Electron, Inc., Louisville, CO, USA, 80027) was performed using 5 µl of serum following the manufacturer's instructions.

Flack, C.P. and J.W. Woollen. 1984. Prevention of interference by dextran with biuret-type assay of serum proteins. Clinical Chemistry. 30:559-661.

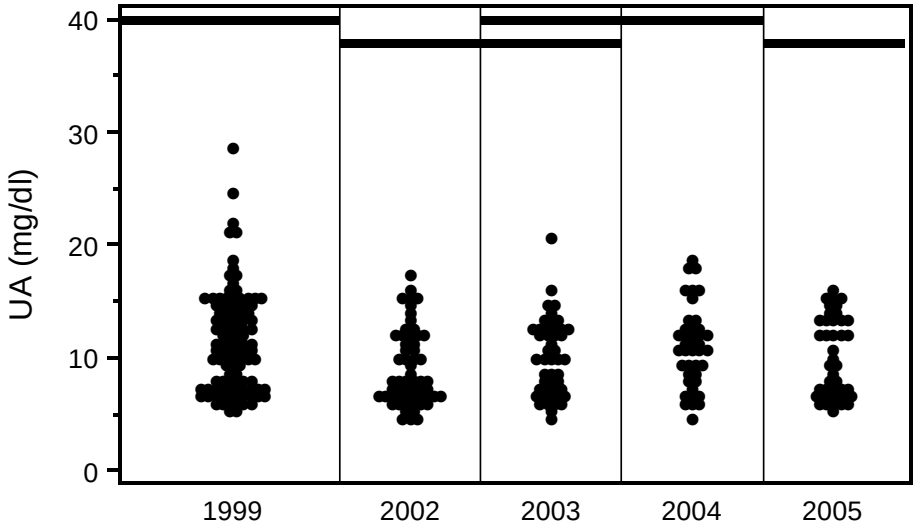
Figure A9. Distribution of albumin (ALB) concentrations as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method. The bromocresol-green reaction using the method of Doumas et al (1972; Cat. No. 1105-250, Thermo-Electron, Inc., Louisville, CO, USA, 80027) was performed using 10 µl of serum following the manufacturer's instructions.

Doumas, B.T., R.L. Arends, and P.C. Pinto. 1972. Determination of albumin in serum. Standard Methods of Clinical Chemistry 7:175-189.

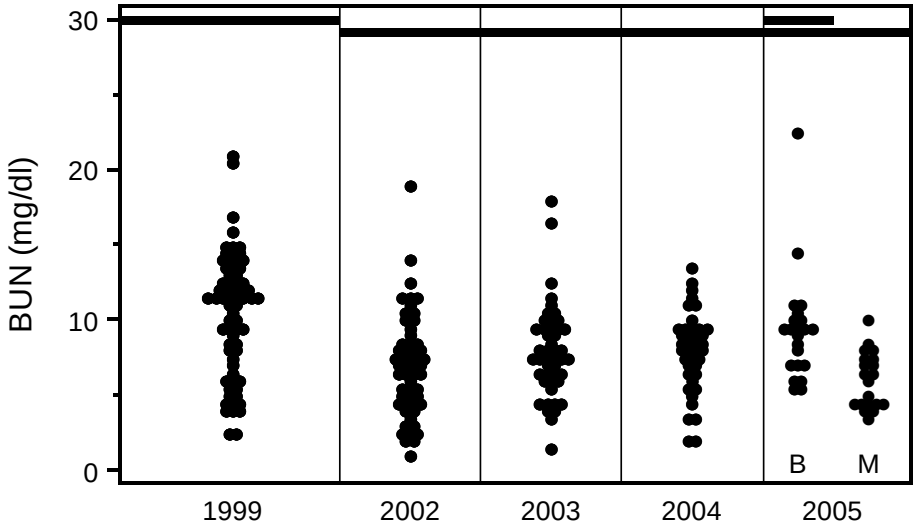
Figure A10. Distribution of uric acid (UA) concentrations as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The production of quioneimine dye by sequential uricase and peroxidase reactions following the methods of Trinder (1949), Kabasakalian et al (1973), and Trivedi et al. (1976; Cat No. TR 24321, Thermo-Electron, Inc., Louisville, CO, USA, 80027) was performed using 2 µl of serum following the manufacturer's instructions.

Kabasakalian, P., S. Kalliney, and A. Wescott. 1973. Determination of uric acid in serum, with use of uricase and a tribromophenol—aminoantipyrine chromogen. Clinical Chemistry 19:522-524.

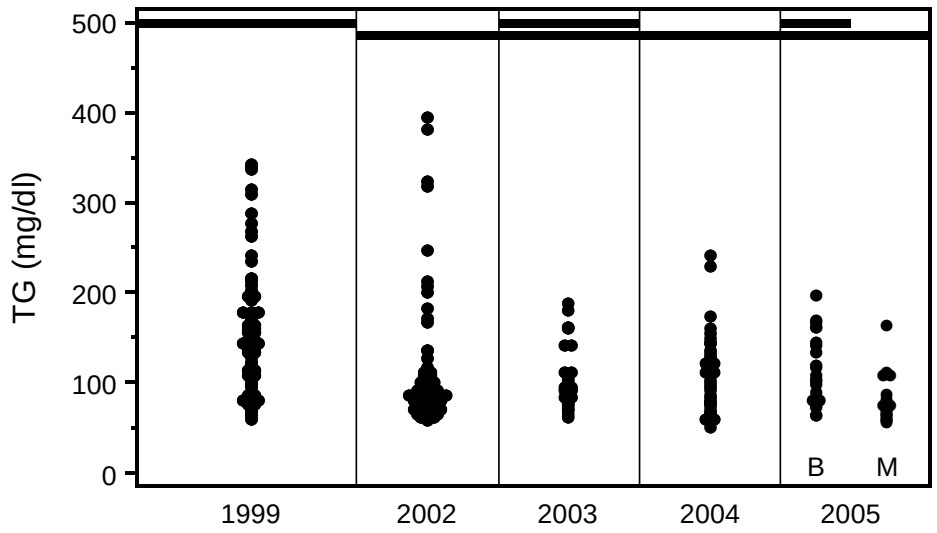
Figure A11. Distribution of blood urea nitrogen (BUN) concentrations as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The production of oxidized nicotinamide dinucleotide (NAD) the sequential urease and glutamate dehydrogenase reactions, following the method of Tiffany et al. (1972; Cat No. TR 12421, Thermo-Electron, Inc., Louisville, CO, USA, 80027) was performed using 10 µl of serum following the manufacturer's instructions.

Tiffany, T.O., J.M. Jansen, C.A. Butris, J.B. Overton, and C.D. Scott. 1972. Enzymatic kinetic rate and end-point analyses of substrate, by use of a GEMSAC fast analyzer. Clinical Chemistry 18:829-840.

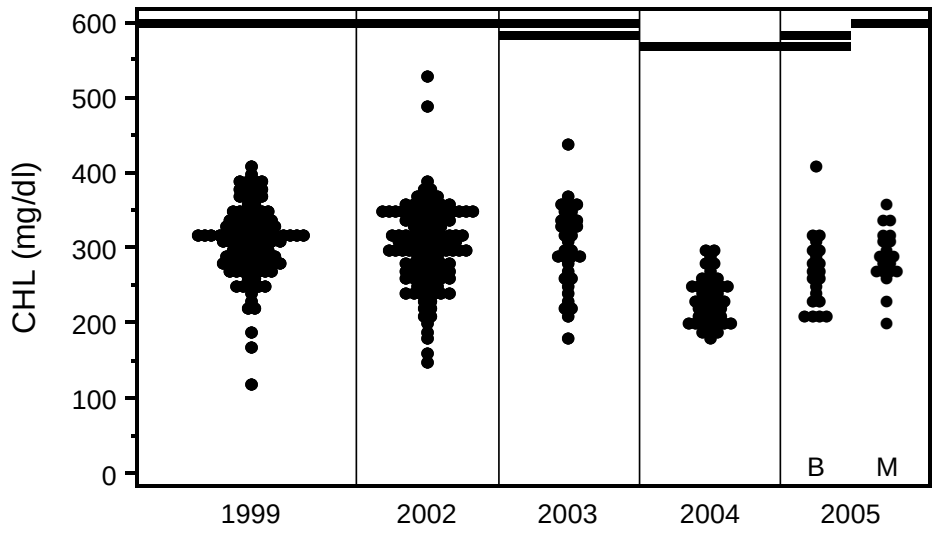
Figure A12. Distributions of triglycerides (TG) concentrations as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The production of quioneimine dye by a sequence of enzymatic reactions, following the method of McGowan et al. (1983), was performed using 2 µl of serum following the manufacturer's instructions (Cat No. 2780-250, Thermo-Electron, Inc., Louisville, CO, USA, 80027).

McGowan, M.W. J.D. Artiss, D.R. Strandbergh, and B. Zak. 1983. A peroxidase-coupled method for the colorimetric determination of serum triglycerides. Clinical Chemistry 29:538-542.

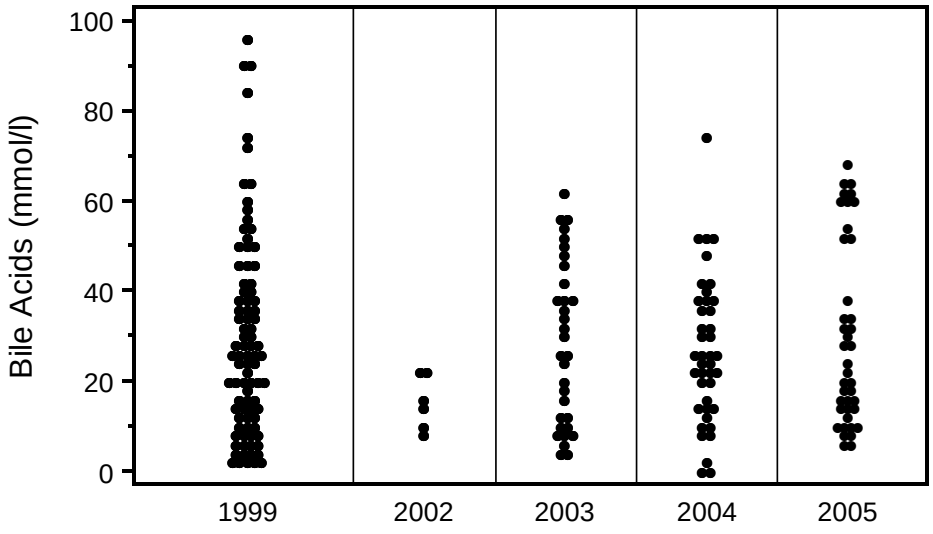
Figure A13. Distributions of cholesterol (CHL) concentrations as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The reaction producing a quioneimine dye by a sequence of enzymatic reactions, following the method of Allain et al. (1974) was performed using 2 µl of serum, following the manufacturer's instructions (Cat. No. 234-60, Diagnostic Chemical Limited, Charlottetown, Prince Edward Island, Canada, C1E 2A6).

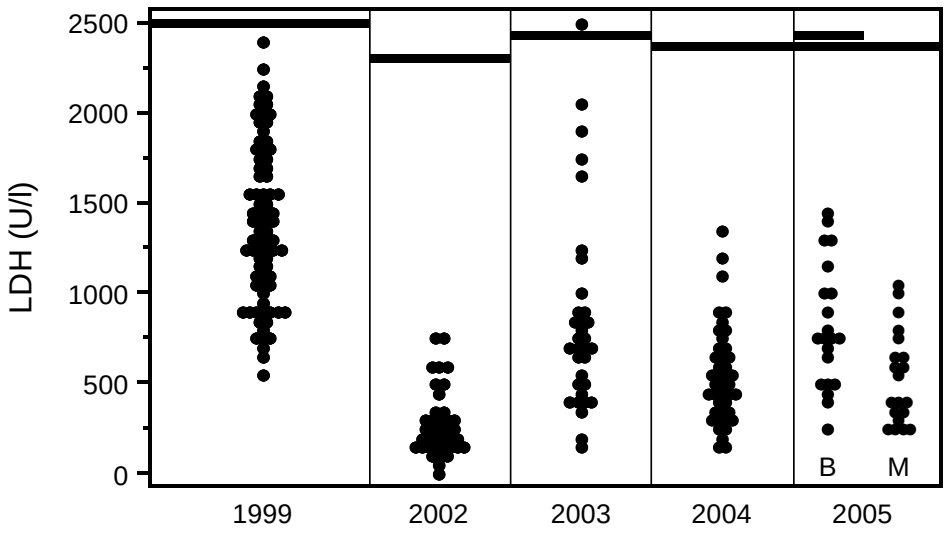
Allain, C.C., L. Poon, S.G. Chan, W. Richmond, P. Fu. 1974. Enzymatic determination of total serum cholesterol. Clinical Chemistry 20:470.

Figure A14. Distributions of bile acids (BA) concentrations as a function of year.



Analytical method: The reaction producing thio-NAD by enzymatic recycling was performed using 2 µl of serum, following the manufacturer's instructions (Cat. No. DZ042A-K01, Diazyme Inc., San Diego, CA, USA 92121).

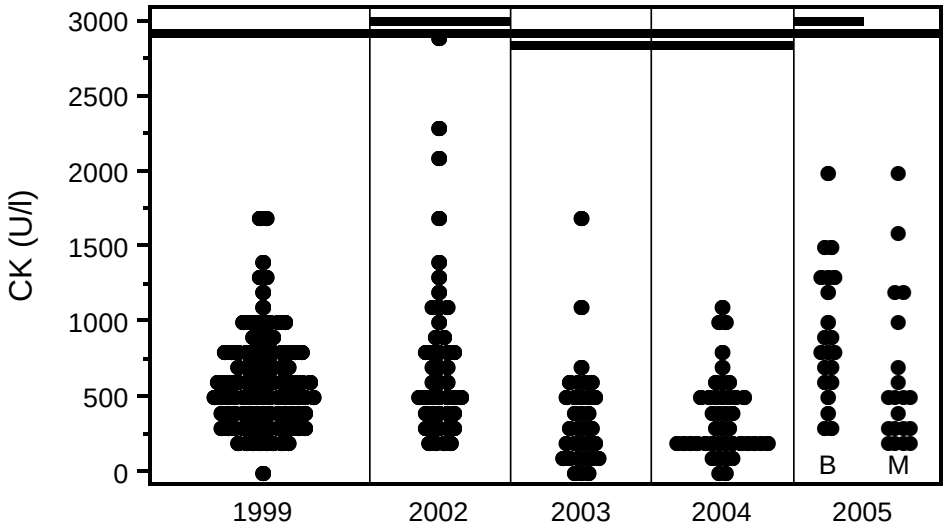
Figure A15. Distribution of lactate dehydrogenase (LDH) activities as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The reaction producing reduced NADH by this enzyme, following the method of the Bais and Philcox (1994), was performed using 5 µl of serum, following the manufacturer's instructions (Cat. No. 324-10, Diagnostic Chemical Limited, Charlottetown, Prince Edward Island, Canada, C1E 2A6).

Bais, R and M. Philcox. 1994. Approved recommendation on IFCC methods for measurement of catalytic concnetration of enzymes. Part 8. IFCC method for lactate dehydrogenase (L-lactate: NAD^+oxidoreductase, EC 1.1.1.27. International Federation of Clinical Chemistry (IFCC). European Journal of Clinical Chemistry and Clinical Biochemistry 32:639-655.

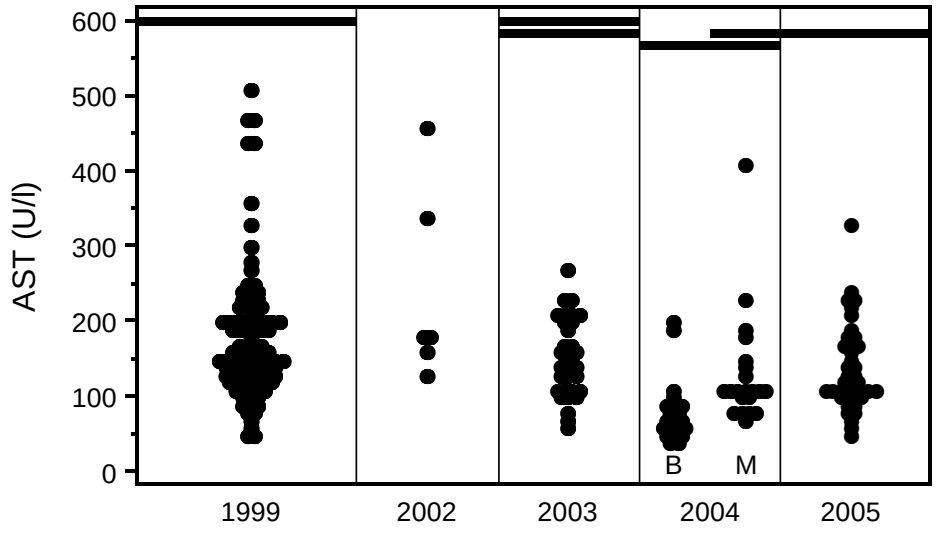
Figure A16. Distributions of creatine kinase (CK) activities as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The production of reduced NADPH by a sequence of enzymatic reactions was measured using 5 µl of serum, following the manufacturer's instructions (Cat. No. 326-10, Diagnostic Chemical Limited, Charlottetown, Prince Edward Island, Canada, C1E 2A6).

Recommendation of the German Chemical Society Standardization of Methods for the Estimation of Enzymes Activities in Biological Fluids. 1977. Journal of Clinical Chemistry and Clinical Biochemistry 15:225-260.

Figure A17. Distributions of aspartate aminotransferase (AST) activities as a function of year. Lines at top connect years with similar means, i.e. not significantly different.



Analytical method: The production of oxidized NAD by a sequence of enzymatic reactions was performed using 10 µl of serum following the manufacturer's instructions (Cat No. TR70121, Thermo-Electron, Inc., Louisville, CO, USA, 80027).

International Federation for Clinical Chemistry. 1986. IFCC Method for L-aspartate aminotransferase. Journal of Clinical Chemistry and Clinical Biochemistry 24:497-510.