Vol. 54: 409–426, 2024
https://doi.org/10.3354/esr01351

ENDANGERED SPECIES RESEARCH
Endang Species Res

Published August 15

Check for updates

OPEN ACCESS CC BY

Combining reproductive endocrinology and ROC analysis to identify changes with sex, age, and pregnancy status in botos *Inia geoffrensis*

Todd R. Robeck[1,*], Rodrigo S. Amaral[2], Karen J. Steinman[3], Jason L. Blum[4,8], Anthony R. Martin[5], Gisele A. Montano[1,3], Don R. Bergfelt[6], Vera M. F. da Silva[7]

[1]SeaWorld Florida, Orlando, FL 32821, USA
[2]Federal Institute of Education, Science and Technology of the Amazonas–IFAM/CMZL, Av. Cosme Ferreira 8045, Manaus 69086-475, Brazil
[3]Species Preservation Laboratory, SeaWorld California, 2595 Ingraham Rd, San Diego, CA 92109, USA
[4]Department of Environmental Medicine, New York University School of Medicine, Tuxedo, NY 10987, USA
[5]Centre for Remote Environments, University of Dundee, Nethergate, Dundee DD1 4HN, UK
[6]Department of Biomedical Sciences, Ross University School of Veterinary Medicine, Basseterre KN 00265, Saint Kitts and Nevis, West Indies
[7]Laboratory of Aquatic Mammals, National Institute of Amazonian Research–INPA, Av André Araújo 2936, Manaus 69067-375, Brazil
[8]*Present address:* Product Safety Labs, 2394 US Highway 130, Dayton, NJ 08810, USA

ABSTRACT: Reproductive hormone profiles were described against physical characteristics during growth and development in male and female botos (Amazon River dolphins *Inia geoffrensis*) and during pregnancy. We determined hormone concentrations in 226 wild botos located in Mamirauá Reserve, Amazonas State, Brazil, as a part of the long-term population monitoring program known as Projeto Boto. Additionally, we applied receiver operator characteristic (ROC) analysis to compare diagnostic probabilities of using ultrasound, hormones, or combinations of these to detect pregnancy. Based on single-point analysis of serum testosterone (T), males with <2.5 ng ml^{-1} T and a mean 163 cm total body length were classified as immature, 2.5 to <4.9 ng ml^{-1} T and 183 cm as pubescent, and >5 ng ml^{-1} T and 227 cm as adult botos. For females, only progesterone (P4), T, relaxin (Rlx), and the combination of P4 × T^2 were significantly different between non-pregnant and pregnant females, but androstenedione (A4) and the P4:T ratio were not. ROC analysis indicated that ultrasound and P4 × T^2 were considered excellent as pregnancy diagnostic tests, and P4, T, and Rlx were classified as good predictors. Results indicated that negative and positive predictive probabilities from each diagnostic test could be used to accurately predict a pregnancy and calf loss rate of 13% for this population. Application of these methods for evaluating wild population reproductive success from a single serum sample can now be used for health evaluations of wild populations of boto and provide timely information for the development or evaluation of any conservation initiatives.

KEY WORDS: Receiver operator characteristic analysis · ROC · Pregnancy diagnosis · Boto · Testosterone · Androstenedione · Androgens · Relaxin · Progesterone · Amazon river dolphin

1. INTRODUCTION

The Amazon River dolphin *Inia geoffrensis*, or boto, from the suborder of odontocetes of Cetacea, was listed on the International Union for Conservation of Nature (IUCN) Red List as Vulnerable in 1996. In 2008, this classification was changed to Data Deficient due to limited available information on threats,

*Corresponding author: todd.robeck@unitedparks.com

© The authors 2024. Open Access under Creative Commons by Attribution Licence. Use, distribution and reproduction are unrestricted. Authors and original publication must be credited.

Publisher: Inter-Research · www.int-res.com

ecology, and population numbers and trends. This classification is a broad category that encompasses species considered critically endangered, endangered, and vulnerable. Thereafter, our knowledge in these areas grew, and in 2018, the species was listed on the IUCN Red List as Endangered (da Silva et al. 2018a,b). Recent evidence suggests that threats to the population are increasing (Martin & da Silva 2022). These threats should incentivize wildlife managers and governmental agencies to develop management plans and regulations designed to decrease the mortality of individuals. However, in addition to decreasing mortalities, population recruitment, and external factors that effect its success, should not be overlooked as primary drivers for sustaining long-lived endangered species (Manlik 2019). Therefore, to provide conservation biologists with the tools necessary to develop management programs for the Endangered boto, additional studies are required to further our understanding of their basic reproductive physiology and to develop tools for evaluating the reproductive health of individual botos.

Botos are confined to freshwater regions of the Amazon and Orinoco River basins of South America (da Silva et al. 2018a). In 1994, Projeto Boto began as a long-term study of river dolphins in the Mamirauá Reserve of Brazil. Since then, approximately 650 individually identifiable botos have been examined by routine observations and periodic physical examinations, resulting in numerous publications that have encompassed habitat (da Silva & Martin 2000, Martin & da Silva 2004, Martin et al. 2004, da Silva et al. 2018a) and general biological characteristics (Martin & da Silva 2006, 2018, de Mello & da Silva 2019, Robeck et al. 2019). More recent studies involving the same population of boto have characterized various reproductive parameters (e.g. pregnancy rates, gestation length, calf birth weight) in adult females and calves (Martin & da Silva 2018) as well as circulating concentrations of thyroid hormones in association with sex, age, and pregnancy status (Robeck et al. 2019). Notably, the latter has provided a comprehensive set of thyroid hormone data from a healthy population of wild boto to establish reference intervals. These data will contribute to future research directed at assessing thyroid function and evaluating the health status of individuals or populations. Given that basic knowledge and understanding of reproductive endocrinology in the boto is limited, additional studies are required to develop corresponding reference intervals as an additional aid to assess reproductive function in this wild population.

Knowledge of basic endocrinology of various reproductive states in adult female boto can also provide alternative methods for pregnancy diagnosis in these free-ranging animals. Although real-time ultrasonography is currently the most definitive tool to diagnose pregnancy in non-domestic or wildlife species (Hildebrandt et al. 2003), there can be logistical and technical limitations, especially when working under remote field conditions in an aquatic environment. Immuno-analysis of circulating concentrations of progesterone (P4) is generally considered the 'gold standard' for hormonal detection of pregnancy in cetaceans (O'Brien & Robeck 2012). As applied in common bottlenose dolphins *Tursiops truncatus*, relatively high P4 concentrations can be diagnostic of pregnancy status; however, because an elevation in P4 in a single sample may not necessarily be pregnancy-specific, repeated blood samples with consistently high P4 are needed for confirmation. Although this approach is possible for aquarium-based animals, repeated collection of blood samples in free-ranging animals is not readily feasible (Trego et al. 2013). Through a series of recent studies of bottlenose dolphins (Bergfelt et al. 2011, Bergfelt et al. 2017) and killer whales *Orcinus orca* (Robeck et al. 2018), circulating concentrations of relaxin (Rlx) have been characterized in relation to reproductive status. In general, relatively high concentrations during the second and third trimesters of pregnancy compared to low concentrations associated with the estrous cycle, pregnancy loss, and during the early postpartum period are indicative of Rlx as pregnancy-specific. In this regard, the use of increased Rlx concentrations as a diagnostic indicator of pregnancy has been evaluated in free-ranging bottlenose dolphins (Bergfelt et al. 2014). Correspondingly, the concordance between hormone-diagnosed pregnancies and known pregnancies based on 9 observed cow–calf pairs was 78% for Rlx and 100% for P4 (Bergfelt et al. 2014). Hence, the combination of Rlx with P4 in a single-sample analysis is more informative and can potentially enhance the accuracy of hormonal diagnosis of pregnancy in free-ranging bottlenose dolphins and, perhaps, other wild cetaceans.

Recent longitudinal studies in aquarium-based female bottlenose dolphins and killer whales have characterized circulating concentrations of testosterone (T) and androstenedione (A4) throughout pregnancy (Steinman et al. 2016, Robeck et al. 2017). In general, both androgens progressively increase by about the second trimester, reach maximum concentrations during the third trimester, and dramatically decline at parturition. Recent evidence suggests that

in these odontocete species, P4 is primarily produced by the corpus luteum and androgens primarily from the placenta (Robeck et al. 2017, 2018, 2021, Legacki et al. 2020, Conley et al. 2021), and Rlx likely from both corpora luteal and placenta (Bergfelt et al. 2011, Robeck et al. 2018). Given that the secretion of progestagens, androgens, and Rlx primarily relies on the placenta, a combination of these hormones in single-sample analysis may not only be diagnostic of pregnancy but prognostic of fetal and placental health and, therefore, predictive of pregnancy outcomes (Robeck et al. 2018).

Receiver operating characteristic (ROC) curve analysis is a statistical approach based on sensitivity and specificity (positive and false positive) of continuous data against a binary outcome (e.g. non-pregnant or pregnant) that can be used to develop cutoff points within continuous data from which diagnostic probabilities for a binary outcome can then be assessed (Faustini et al. 2007, Hart 2016). In clinical practice, diagnostic probabilities for the presence or absence of disease are used within a decision tree to determine the need for future tests or treatments (Saeed et al. 2015, Choe et al. 2023). For reproductive exams, cut-off or threshold values of various hormones allow for the development of probabilities that an individual is either pregnant or not pregnant or, within a pregnancy, are used to assess the probabilities of carrying the fetus to term (Al-Sebai et al. 1995, Faustini et al. 2007, Abdullah et al. 2014). However, a potential application of this technique has yet to be used to evaluate or compare the relative predictive ability of various hormone concentrations for pregnancy detection in cetaceans.

The objectives of the present study for a free-ranging population of boto in Central Amazon were to (1) analyze serum concentrations of P4, T, A4, and Rlx in relation to sex, age, and reproductive status; (2) develop reference ranges for each hormone relative to sex, age, and stage of pregnancy; and (3) apply ROC statistical methods to evaluate the application of one or more hormones alone or in combination to be diagnostic and prognostic of pregnancy status and, based on subsequently observed cow–calf pairs, reproductive failure rates.

2. MATERIALS AND METHODS

2.1. Animals and animal classifications

A total of 226 botos *Inia geoffrensis* were evaluated for general health from 2003 through 2015 during an annual Projeto Boto capture–release–recapture campaign at the Mamirauá Sustainable Development Reserve, Brazilian Amazon (da Silva & Martin 2000). Details of the animal capture and release techniques have been previously described (da Silva & Martin 2000). Briefly, the study area whereby the capture and releases occur includes a channel with an entrance to the Mamirauá Lake and channel complexes known as the Mamirauá System. At the end of the dry season, narrow channels connecting the Mamirauá Lake to the river systems become important routes for botos. The movements through these channels peak near the end of the dry season, and it is during this time that a field expedition (usually 20 d in length) is conducted. All the animals that pass through the channel are captured, examined, and then released (da Silva & Martin 2000). Subsequent to capture and physical restraint as previously described (de Mello & da Silva 2019, Robeck et al. 2019), blood samples were collected, and total body length (TL; tip of rostrum to fluke notch), sex, and pregnancy status were determined prior to release. For females, classification of adult and immature animals was based on pregnancy status, body length (adults: >179 cm), and P4 serum concentrations of >1 ng ml^{-1}, the threshold concentration that is presumptively indicative of ovulation (Martin & da Silva 2018). For males, botos were initially classified based on TL into immature (calves and juveniles: <188 cm TL) or adults (>187 cm TL) as previously described (Robeck et al. 2019) and then further refined based on serum T concentrations as determined herein.

Pregnancy status (pregnant, non-pregnant) was diagnosed using an ultrasound unit (Sonosite 180 Plus, Sonosite) in conjunction with a 2–5 MHz multi-frequency transducer (Convex Array 180 Plus/ Elite-C60, Sonosite). Pregnancy confirmation was based on the imaging of intrauterine fluid and conceptus as described for odontocetes (Lacave et al. 2004, Robeck et al. 2015); non-pregnancy status was based on the absence of imaging any of these aspects. Based on a 12 mo gestation period (Martin & da Silva 2018), stages of pregnancy were classified as early (months 0–4 or days 0–112), mid (months 5–8 or days 113–224), and late (months 9–12 or days 225–333). Determination of pregnancy stage at the time of sample collection was done using 2 methods and based, in part, on the assumption that gestation lengths and periparturient calf sizes are similar between boto and bottlenose dolphins (da Silva et al. 2023). The first method was to apply bottlenose dolphin fetal growth models (Lacave et al. 2004) based on thoracic or

biparietal diameter (cm) towards comparable boto ultrasonographically determined fetal measurements to estimate fetal age and days until gestational term. From these data, we then approximated the stage of pregnancy of the animal at the time of ultrasonographic exam and sample collection. The second method utilized for pregnancy stage estimation occurred when a serum sample was collected without an ultrasound exam; the female was determined to be pregnant based on increased serum P4 and observation of a neonatal calf during the subsequent surveys. By estimating the age of the calf coupled with the estimated gestation period, we could retrospectively estimate the stage of pregnancy for that sample time point. Non-pregnant, lactating adult females with or without an accompanying calf were initially categorized as lactating.

2.2. Photo-identification of cow–calf pairs

The techniques for photo-identification, boat-based survey methods, and data collection have been previously described (Mazzoil et al. 2005, Speakman 2006, Urian et al. 2015). Although documentation of reproductive success using photo-identification has substantial methodological limitations, the calf of a cow–calf pair is typically characterized by being 50–75% the length of the cow and in close proximity (Urian & Wells 1996, Gaona Calderón et al. 2018). For the present study, photo-identification records were searched for the sighting of cow–calf pairs that corresponded to any of the adult female boto evaluated herein. Sightings of cow–calf pairs were used conservatively as an indication of observed or known pregnancies that may be in concordance with the hormonal diagnosis of pregnancies based on maternal blood samples collected within 12 mo of a previous health assessment. Females sighted without calves were not used as non-pregnant because of the periodicity of surveys, unrecognized pre- and post-natal losses, predation of calves, and other events that may occur prior to sightings, all of which preclude a definitive diagnosis of non-pregnancy (Urian & Wells 1996, Wells et al. 2014, Urian et al. 2015).

2.3. Collection of blood

Blood samples were collected from the ventral tail fluke using a 19-gauge winged blood collection set. Whole blood was collected into BD Vacutainers (Becton Dickenson) containing activated thrombin and allowed to clot for 20 min at ambient temperature. Then, samples were moved to a refrigerator at 4°C for holding until centrifugation. Samples were centrifuged at 1000 × *g* for 15 min within 4 h of collection and then stored at −20°C until analysis. Blood serum samples were shipped frozen at −150°C in a dry shipper from the National Institute of Amazonian Research in Manaus, Brazil, to the SeaWorld & Busch Gardens Species Preservation Laboratory located in San Diego, California, USA. Upon arrival, samples were stored at −80°C until hormone analysis.

2.4. Androgen analysis (serum extractions)

Solvent extraction of boto serum was conducted prior to the analysis of T and A4 as previously described (Robeck et al. 2017). Serum (0.15 ml) and extraction buffer (0.15 ml, 0.2 M phosphate-buffered saline, pH 7.5) were added to a 15 ml polypropylene tube (Falcon, Becton Dickenson). Thereafter, 3 ml of diethyl ether (Acros Organics, ThermoFisher) was added to all tubes. The tubes were capped and placed on a multitube vortexer for 5 min (1300 rpm; BenchMixer, Benchmark Scientific). After vortexing, samples remained on the benchtop for 10 min to allow for separation of the ether layer. Upon layer separation, samples were placed in a −80°C freezer for 20 min to allow the non-solvent layer to freeze. Under a fume hood, the supernatant was rapidly poured off into 16 × 100 mm glass tubes (ThermoFisher) before the non-solvent layer was thawed and evaporated under a stream of compressed nitrogen gas (20–30 min total drying time). Dried extracts were reconstituted in 0.3 ml of extraction buffer and stored at −20°C until analysis.

Due to an insufficient sample volume of boto sera, evaluation of extraction efficiency was not conducted. However, based on previous research on killer whales (Robeck et al. 2017), a single extraction step was sufficient to recover ~90% of respective hormones. Again, because of the insufficient volume of boto sera, to approximate the efficiency of each sample's extraction batch, a pool of female bottlenose dolphin sera was used to assess spike and recovery for each sample extraction batch. An unspiked and spiked sample of the extraction control was prepared in the same manner as described above, and a second sample containing 0.15 ml of extraction control and 0.15 ml of a spiked A4 standard (20 pg A4 per well solution in extraction buffer) was combined and extracted as described above. The reconstituted control and spiked control were then assayed, and the recovery of the spiked A4 control was determined to

yield the extraction efficiency for each batch. Any extraction batch displaying a control recovery of <85% was repeated to ensure adequate efficiency of the procedure and uniformity across extraction batches.

2.5. Testosterone (T) assay

T concentrations were measured using a single antibody, direct enzyme immunoassay (EIA) on aliquots (0.010–0.025 ml, depending on the sample concentration) of the reconstituted extracts as previously described in detail for use in the bottlenose dolphin (Steinman et al. 2016).

Cross-reactivities of the T polyclonal (rabbit) antibody (R156/7; C. Munro, UC Davis, CA, USA) are 100% T, 57% dihydrotestosterone, 0.37% A4, and <0.05% with other tested analytes (Dloniak et al. 2004). Assay sensitivity was 46 pg ml^{-1}, and intra-assay variation was <10%. Additionally, standards were run in triplicate on the microtiter plate, in duplicate at the beginning of the plate, and in singlet at the end, and the intra-assay CV was also <10% for all standards within and between each assay. Parallel displacement of serum compared to the standard curve was demonstrated ($r = 0.993$, $p < 0.05$), and the recovery of known concentrations of standard to serum was 96.3 ± 12.2% (linear regression, $y = 0.7x + 1.36$, $r^2 = 0.992$, $F_{1,7} = 851.06$, $p < 0.05$). Inter-assay CVs for 2 quality controls with antibody binding at approximately 30 and 70% were 4.0 and 4.9%, respectively (n = 9).

2.6. Androstenedione (A4) assay

A4 concentrations were measured using an EIA kit (40-056-205044, GenWay Biotech) as previously described for killer whales (Robeck et al. 2017). Aliquots (0.010–0.025 ml, depending on the sample concentration) of the reconstituted extracts were processed according to the kit instructions. Cross-reactivities of the A4 polyclonal antibody utilized are 100% A4, 0.1% dehydroepiandrosterone, and 0.1% T; all other steroids tested were <0.01%. Parallel displacement of serum compared to the standard curve was demonstrated ($r = 0.998$, $p < 0.05$). The recovery of known concentrations of standard to extracted serum was 113.9 ± 3.7% (linear regression, $y = 1.16x - 0.22$, $r^2 = 0.999$, $F_{1,3} = 4397.01$, $p < 0.05$), thereby demonstrating negligible matrix interference in the EIA. Assay sensitivity was 19.0 pg ml^{-1} and intra-assay CV was <10%. The inter-assay CV for a high and low control with antibody binding at approximately 20 and 65% were 4.1 and 8.3%, respectively (n = 7). The CV for a pool of serum used as a control to measure the recovery and efficiency of the extraction procedure was 4.1% and the mean recovery rate was 91.3 ± 3.7% (n = 12).

2.7 Progesterone (P4) assay

Immuno-analysis of serum P4 was measured on non-extracted serum (0.050–0.0025 ml, depending on the sample concentration) by a direct EIA (Munro & Lasley 1988, Graham et al. 2002) using the antibody CL425 (1:8000) with the respective horseradish peroxidase-conjugated hormone (C. Munro, UC Davis, CA, USA). P4 cross reactivities with the antibody were 4-pregnen-3α-ol-20-one, 188%; 4-pregnen-3β-ol-20-one, 172%; 4-pregnen-11α-ol-3,20-dione, 147%; 4-pregnen-3,20-dione (P4), 100%; 5α-pregnen-11-3β-ol-20-one, 94%; 5α-pregnan-3α-ol-20-one, 64%; 5α-pregnane-3,20-dione, 55%; 5β-pregnane-3β-ol-20-one, 12.5%; and other steroids, <0.1% (Graham et al. 2002). Parallel displacement of serum compared to the standard curve was demonstrated ($r = 0.995$ [Graham et al. 2001], $p < 0.05$). Assay sensitivity was 0.08 ng ml^{-1} and intra- and inter-assay CVs were <10%.

2.8. Relaxin (Rlx) assay

Analysis of serum immunoreactive Rlx was conducted using a validated radioimmunoassay (RIA) developed for use in bottlenose dolphins and killer whales as previously described (Bergfelt et al. 2011, Robeck et al. 2018). The main components of the assay consisted of H2 human Rlx as 125I-labeled ligand (30 000 cpm $tube^{-1}$), rabbit anti-porcine Rlx R6 as the primary antibody (working dilution, 1:40 000), and synthetic canine Rlx as a reference standard (0.2–50 ng $tube^{-1}$). Samples were analyzed in duplicate, with results reported as average concentrations (ng ml^{-1}) relative to the Rlx standard curve. Assay sensitivity was 14.1 ng ml^{-1} and intra-assay CV was 10.1% with pooled serum from several pregnant bottlenose dolphins.

2.9. Statistical analyses

Prior to any analysis, data were separated by sex. All data were analyzed using STATA® (v.18; Stata Corp.). For females, we analyzed the hormone data using a mixed effect restricted maximum likelihood

(REML) regression model (West et al. 2014) to quantify the relationship between the dependent variables (hormone concentrations) and fixed affect variables, with animal ID as the random intercept variable with an unstructured covariance. For all linear mixed models (LMMs), degrees of freedom were estimated using the Kenward-Roger approximation for small sample sizes. The 2 fixed effect categorical variables included reproductive status (non-pregnant, pregnant, non-pregnant lactating [NPL]) and age (adult, immature). A separate analysis was performed with both fixed effect variables for each dependent variable hormone, P4, T, A4, and Rlx. Additionally, the ratio of P4:T, which has been demonstrated to change significantly during different stages of pregnancy in killer whales (Robeck et al. 2018), was used as a dependent variable and analyzed as described above.

All final mixed models were checked for normality using quantile plots of the standard residuals. If quantile–quantile plots of standardized residuals exhibited non-normal distribution, raw data were transformed as indicated by the Shapiro-Wilk test until model residuals were normalized. Pairwise comparisons of estimated marginal means were conducted using the margins command with Šidák correction ($p \leqslant 0.05$). Unless specified, data are expressed as back-transformed means ± SE. Unless stated, statistical significance was considered at $p \leqslant 0.05$.

For males, a 2-sample *t*-test with unequal variance using Welch's approximation for degrees of freedom was used for separate comparisons of mean T and A4 concentrations between ages (adult, immature). For comparisons between immature, pubescent, and adult concentrations of hormones, we used a 1-way ANOVA and a post hoc multiple pairwise comparisons test using Šidák correction.

Finally, reference intervals were created for the central 95% range (2.5th, 25th, 50th, 75th, and 97.5th percentiles) for each hormone concentration partitioned within each sex, age group, and physiologic state (male: immature, pubescent, adult; female: juvenile female, adult female, pregnant female) and presented in tabular form. Reference intervals were calculated based on the recommendations from the Clinical and Laboratory Standards Institute (Clinical and Laboratory Standards Institute 2008) using a nonparametric bootstrap-based procedure (Linnet 2000, Nabi et al. 2019). Using bootstrap resampling techniques (1000 reps), the mean 2.5th and 97.5th percentiles from the 1000 replicates of each group were determined along with the 95% bootstrapped confidence intervals (CIs) for each mean, respectively. Bootstrapped 95% CIs were only determined for groups with samples from a minimum of 20 individuals. For analytes that had values that represented fewer than 20 individuals, only the range of these values was reported (Linnet 2000). Pearson pairwise correlations analysis was used within each sex to evaluate relationships between hormone concentrations to characterize potential physiologic connections.

2.10. Diagnostic testing and ROC analysis

In order to determine the accuracy of using one or more of the hormones (P4, T, A4, Rlx) or a combination of hormones (e.g. P4:T ratio) to predict whether an animal was pregnant or not, we used nonparametric ROC analysis to determine the total area under the ROC curves (tAUC) for each test as compared against a gold standard of 'true pregnancy' (Pepe et al. 2009). Additionally, we also determined partial AUC (pAUC) fixed at <10% false positive rate (FPR). The use of pAUC may provide an improved, clinically relevant evaluation of the predictive value of these hormones as a single-sample pregnancy test (Pepe et al. 2009). To develop a gold standard for the ROC curve, we relied on ultrasonography to indicate whether the animal was pregnant or not pregnant. An ultrasound exam is considered the gold standard for confirming pregnancy in cetaceans (Ivančić et al. 2020). However, for our gold standard development of true positive or negative, we had to account for possible false negatives that are most likely to occur during early gestation when the conceptus is undetectable using transabdominal ultrasonography (Fricke 2002, Robeck et al. 2015). Therefore, any female that did not have an identified conceptus or calf on exam but was later deemed to have been observed with a calf within less than 1 yr of the ultrasound exam was considered to have been pregnant. This combination (ultrasound + calf observation) was then used for the gold standard in subsequent ROC analysis. We first determined which hormones produced the greatest tAUC and pAUC (fixed at 10% FPR) using a multiple comparison test with Bonferroni correction. Thereafter, a Wald statistic based on the differences in their bootstrapped standard errors was used to generate a p-value for comparisons of tAUC for the ROC curves and to indicate what single hormone tests may provide value as a diagnostic test of pregnancy (Pepe et al. 2009). Cross-validation (CrV) with 10 random splits in the original data to generate an estimated mean CrV AUC was performed to theoretically allow one to infer the applicability of the ROC models toward larger or unknown data sets (Luque-Fernandez et al. 2019). Finally, we wanted to evaluate whether a regression model combining the

top 3 hormones would result in an incremental increase in ROC accuracy and then compare it against those from the individual or combined hormones (T and P4). This was accomplished by a simple logit regression against the gold standard results with P4, and then with T and Rlx added incrementally as covariates in 2 additional models (Janes et al. 2009, Moons et al. 2012). The models were as follows: (1) logit gold P4; (2) logit gold P4 + T; and finally, (3) logit gold P4 + T + Rlx. Predicted values were produced after each regression and then used to compare ROC results from each model as described above.

Optimum cutoff points for each hormone or hormone combination were then determined using the 'roctab' command analysis (STATA). This analysis generated tables of the sensitivity, specificity, and total percent correct at each change in hormone concentration. The 'optimum' cutoff points were then chosen from this table by identifying the minimum hormone concentration that resulted in the greatest percent correct (highest combination of true positive and true negatives) for each hormone within this sample population. These cutoff values were then used to classify all females as non-pregnant or pregnant (0, 1), based on whether the hormone values were at or above (1) or below (0) this threshold. The diagnostic accuracy of these different hormone models as compared to the gold standard could then be determined. Diagnostic accuracy was defined as the positive and negative predictive value as determined from the sensitivity (proportion of pregnant females correctly identified) and specificity (proportion of non-pregnant females correctly identified) at each prevalence point within each animal grouping (ultrasound examined, observed calf groups, or hormone test) and was determined using the 'Diagt' command within STATA (Seed & Tobias 2004, Mandrekar 2010). We then used the results from the diagnostic test analysis to predict the total calf loss (abortion + stillbirths + neonatal and calf deaths prior to 1 yr of age) for the total population as a demonstration of how our results could be used to predict these values within an unknown population.

3. RESULTS

3.1. Animal classifications

Out of a total of 226 botos, 126 were identified as female and 92 were identified as male. Fifteen females were sampled more than once during different seasons, with 2 of these individuals being sampled 3 times for a maximum number of 144 samples. Within these sampling periods, 60 were non-pregnant, 61 pregnant, and 23 NPL. For pregnant and NPL, mean body lengths were 196.3 cm (95% CI: 193.8–198.9 cm) and 200.5 cm (95% CI: 198.1–202.9 cm), respectively. Although the shortest body length of either pregnant or lactating females was 180 cm (range: 180–222 cm), the range in body lengths of the other females classified as sexually mature or adults based on evidence of ovulation (serum P4 >1 ng ml^{-1}) was 162–178 cm. Thus, of the 60 non-pregnant females, 19 were classified as adults, with body lengths of ⩾162 cm but <180 cm, and 14 were classified as immature, with body lengths of <162 cm.

3.2. Stage of pregnancy

Ultrasound determination of fetal size using bottlenose dolphin fetal growth models was used to estimate the stage of pregnancy in 61 pregnant animals (Lacave et al. 2004). For 28 animals within this group, we were able to estimate parturition dates based on calf size, compared these estimates to the predicted stage of pregnancy based on ultrasound data, and found 100% agreement for placing animals within respective stages of pregnancy: early, mid, or late. Therefore, we classified all pregnancies with a high degree of confidence within these 3 categories for analysis. For the analysis, we identified 33 females in the first stage, 19 in the second, and 9 in the third.

3.3. Males

3.3.1. Length at maturation

A 2-sample *t*-test comparison of log T concentrations indicated differences ($t_{59.6}$ = 9.01, p < 0.0001) between immature (<188 cm TL, 2.02 ng ml^{-1}, 95% CI: 1.55–2.65 ng ml^{-1}) and mature males (>187 cm TL, 9.18 ng ml^{-1}, 95% CI: 7.47–11.28 ng ml^{-1}). The scatterplot of serum T versus length provided subjective support for further partitioning based on hormone concentrations and length (Fig. 1). We observed the following: (1) all males >219 cm in length had serum T > 5 ng ml^{-1} (17.4 ng ml^{-1}, 95% CI: 13.8–21.0, range: 5–52.5 ng ml^{-1}, n = 39), which we then used as an indicator of reproductive maturity (i.e. males with serum T > 5 ng ml^{-1} were reproductively mature); (2) the shortest male with serum T > 5 ng ml^{-1} was 174 cm, and multiple males between 174 and 219 cm had serum T concentrations between 4 and 5 ng ml^{-1}. Therefore, based on these hormonal observations in

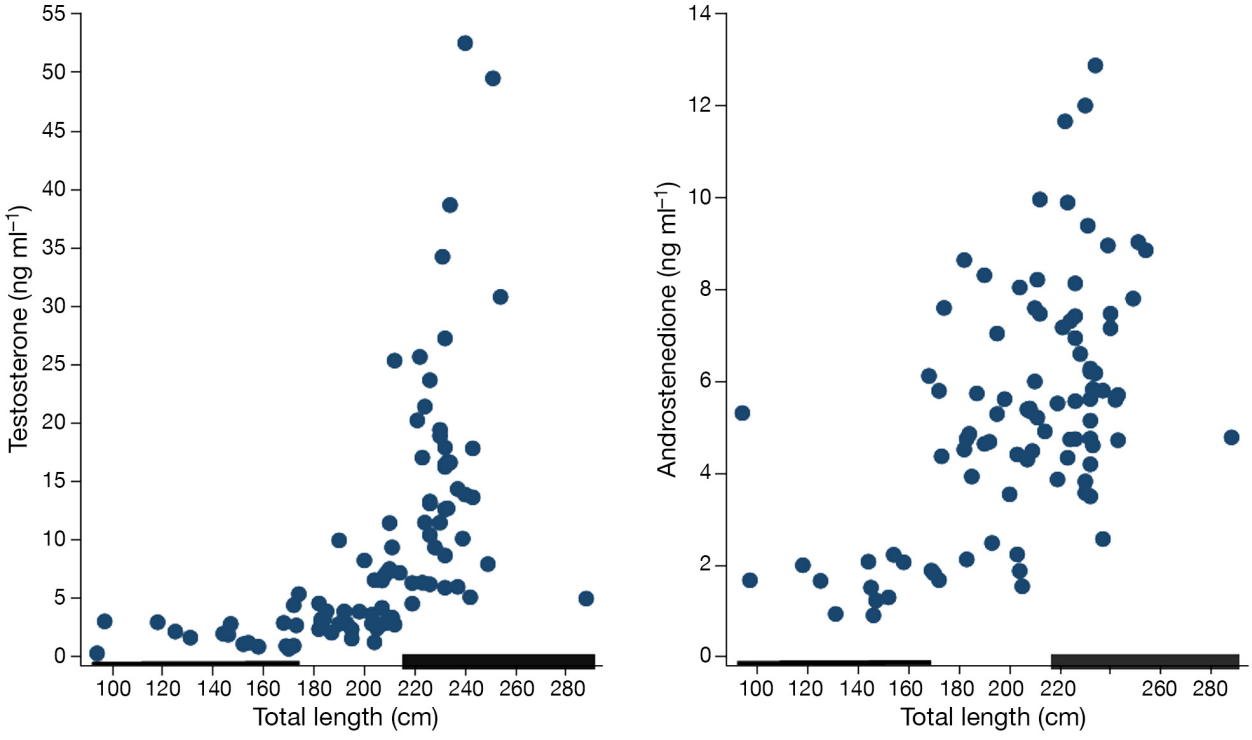

Fig. 1. Scatterplot of serum testosterone and androstenedione versus length in male botos *Inia geoffrensis*. Visual distribution provided subjective support for partitioning reproductive maturity into 3 groups rather than 2 (mature and immature), and based on hormone concentrations at length, we subjectively proposed the following groups: immature (thin black line, total length [TL] < 174 cm), pubescent (non-marked area, TL: 174–219 cm), and adult (thick black line, TL > 219 cm)

combination with body length, we further grouped animals into the following categories: immature (TL < 174 cm), pubescent (TL = 174–219 cm), and adult (TL > 219 cm; Fig. 1). One-way ANOVA analysis of log T concentrations within these groups found differences ($F_{2,89}$ = 140.0, p < 0.0001). Concentrations within each group were as follows: immature, mean 1.73 ng ml^{-1} (95% CI: 1.29–2.3 ng ml^{-1}); pubescent, mean 4.3 ng ml^{-1} (95% CI: 3.5–5.3 ng ml^{-1}); and adult, mean 14.6 ng ml^{-1} (95% CI: 12.0–17.9 ng ml^{-1}) with Sidák post hoc inter-group analysis indicating each were different (p < 0.001) from the other.

3.3.2. Serum A4 concentrations by age group

As expected, 1-way ANOVA of the log A4 concentrations within age groups as defined by serum T was significant ($F_{2,89}$ = 33.00, p < 0.0001). Sidák post hoc comparisons indicated immature (2.33 ng ml^{-1}, 95% CI: 1.89–2.89 ng ml^{-1}) was less (p < 0.003) than pubescent (3.8 ng ml^{-1}, 95% CI: 3.13–4.61 ng ml^{-1}), and both were less than adult concentrations (6.21 ng ml^{-1}, 95% CI: 4.48–7.05 ng ml^{-1}, p < 0.001).

3.3.3. Hormone reference ranges for male botos

Bootstrapped percentiles (2.5th, 50th, 97.5th) for T and A4 by TL are shown in Table 1. Due to the large overlap between T and A4 concentrations, it was clear and unsurprising that TL is an imperfect classifier of reproductive maturity, but these body length guidelines corresponded well to previous work and will help classify animals in the field that could otherwise not be handled for the length of time to collect serum samples or to perform ultrasonographic testicular exams.

3.4. Females

3.4.1. Reproductive hormone concentrations

For some animals, inadequate serum volume was banked, which precluded our ability to analyze all hormones; therefore, some minor differences in total numbers of animals, samples collected, and hormone examined are evident in the results (Table 2). For analysis, all hormones within females were log transformed for the final LMM REML analysis. Compari-

Table 1. Reference intervals (2.5th, 50th, and 97.5th percentiles) and their associated 95% bootstrap (1000 reps) confidence intervals (CI). Serum testosterone and androstenedione reference concentrations within immature (IM; n = 19), pubescent (PM; n = 34), or adult (AD; n = 39) male classes of boto *Inia geoffrensis* in samples collected from 2003 through 2015. For sample sizes of <20, only minimum and maximum analyte values are provided, and thus CIs could not be determined (ND). Total body length (TL; in cm) classes — IM: < 174 cm; PM: 174–219 cm; AD: >219 cm

Hormone (ng ml^{-1})	Age class	2.5th (95% CI)	50th (95% CI)	97.5th (95% CI)
Testosterone	IM	0.3 (ND)	2.0 ± 0.42 (1.3–2.7)	5.4 (ND)
	PM	1.3 ± 0.3 (0.8–1.8)	3.9 ± 0.7 (2.8–5.0)	25.4 ± 7.0 (13.8–37.0)
	AD	5.0 ± 0.3 (4.5–5.5)	14.4 ± 1.6 (11.7–17.0)	53.5 ± 5.3 (43.7–61.3)
Androstenedione	IM	0.9 (ND)	1.9 ± 0.4 (1.2–2.5)	7.6 (ND)
	PM	1.5 ± 0.3 (1.1–2.0)	5.1 ± 0.29 (4.6–5.5)	10.0 ± 0.7 (8.8–11.1)
	AD	2.6 ± 0.5 (1.8–3.4)	6.2 ± 0.5 (5.4–6.9)	12.9 ± 0.7 (11.7–14.0

son of hormone concentrations between adult pregnant and non-pregnant, sexually mature females revealed that only models for P4, T, Rlx, and P4 × T^2 were significant, and the P4:T ratio and A4 were not, with only P4 having significant random effects (Table 2, Table S1 in the Supplement at www.int-res.com/articles/suppl/n054p409_supp.pdf). Post hoc analysis found that P4, T, Rlx, and P4 × T^2 were significantly different compared with non-pregnancy, and A4 and the P4:T ratio were not different (Table 2, Table S1).

LMM REML comparisons of hormones during stages of pregnancy against juvenile and adult non-pregnant females found that all models were significant, with only A4 having significant random effects (Table S2). Post hoc marginal mean comparison found that only A4 was different ($p < 0.05$) for early and non-pregnancy compared to late stage (Table 2). For P4, all pregnancy stages were greater than in non-pregnant females ($p < 0.05$), but not different when compared to each other within pregnancy (Table 2). For both T and P4 × T^2, mid- and late-stage pregnancies were greater than early, which in turn was greater than non-pregnancy. Finally, it should be noted that the concentration of Rlx from one animal was 1710 ng ml^{-1}—twice that of the next highest sample and greater than the upper 2 SDs above the late-stage arithmetic mean (1457.5 ng ml^{-1}). Based on ultraso-

Table 2. Serum marginal mean (95% CI) concentrations of progesterone (P4), testosterone (T), androstenedione, relaxin, and P4 × T^2 combination during pregnancy stages and non-pregnant sexually mature (SM; >161 cm in length) boto females. Number (n) of samples for each group represents the minimum number of samples available for hormone analysis. Not all animals had enough serum volume to run all the hormones. Hormone concentrations were back-transformed after analysis. Šidák correction factor was used for comparisons of marginal means. Only stages that were significantly different ($p < 0.05$) from other groups are listed. *significant difference ($p < 0.001$) between adult females with a calf (pregnant) compared to non-pregnant SM females (2-tailed Student's *t*-test)

Group	Hormone (ng ml^{-1})				
	P4 (n = 126)	T (n = 126)	P4 × T^2 (n = 126)	Androstenedione (n = 126)	Relaxin (n = 117)
Early stage (A) (n = 33)	8.7 (5.9–12.9)	2.41 (1.8–3.2)	51.7 (24.6–108.5)	8.0 (6.4–9.9)	106.7 (78.8–144.4)
Mid stage (B) (n = 17)	10.6 (6.3–17.8)	5.5 (3.7–8.2)	326.8 (119.4–894.7)	11.9 (8.9–15.8)	194.0 (129.9–389.8)
Late stage (C) (n = 8)	9.9 (4.5–21.6)	12.0 (6.7–21.5)	1433.8 (329.4–6240.0)	17.7 (11.6–27.2)	298.9 (163.4–546.8)
Non-pregnant SM (D) (n = 65)	1.4 (1.0–1.8)	0.55 (0.45–0.68)	0.41 (0.24–0.69)	8.8 (7.6–10.2)	18.7 (15.0–23.3)
Šidák group	D < A,B,C	D < A < B,C	D < A < B,C	A,D < C	D < A,B,C; A < C
Pregnant (n = 61)	9.68* (7.42–12.61)	3.59* (2.82–4.55)	128.1* (72.2–227.3)	10.05 (8.55–11.81)	137.8* (108.4–175.3)

nography, this animal was estimated to be in late gestational stage and, based on Rlx concentrations for bottlenose dolphins (Bergfelt et al. 2017), was likely close to term. This animal was observed swimming with a calf 48 d after its health assessment.

3.4.2. ROC analysis for pregnancy prediction and pregnancy outcome

True positive rates (sensitivity) were plotted (ROC curves) versus FPRs (1 − specificity) for serum, P4, T, A4, Rlx, and P4 × T^2 against female pregnancy or non-pregnancy using the gold standard (Fig. 2). The tAUC comparisons indicated that serum P4 (AUC: 0.919) was increased ($p < 0.0001$) compared with the P4:T ratio (AUC: 0.575) and A4 (AUC: 0.582) and was reduced ($p = 0.0029$) when compared against P4 × T^2 (AUC: 0.977). No significant differences were detected in AUC between P4 and Rlx (AUC: 0.925) or T (AUC: 0.934; Table 3). The ROC analysis for the P4:T ratio indicated that it had no predictive value (AUC approaching 0.5) for pregnancy detection and was dropped from further analysis. Although A4 was also of limited value, it was kept in the modeling for context with the other hormones. The mean CrV ROC scores (Table 3) were similar across all hormones and provided some evidence that the predictive value of the results using this assay system could be generalized toward females not used for model development (Luque-Fernandez et al. 2019). Post hoc comparisons of the ROC analysis with the FPR (specificity) held at 10% indicated that no significant differences between any of the hormones were detected, while the combination of P4 × T^2 was significant ($p = 0.05$; Table 3). Cutoff values, sensitivity, specificity, and percent correct are listed in Table 3. Although the hormone combinations of P4 × T^2 provided the largest AUC, we also tested whether a linear combination of the 3 top hormones provided incremental improvements in AUC coverage. Results indicated that a significant incremental increase was observed from the lowest AUC of P4 (AUC: 0.915), to the highest with P4 + T + Rlx ($p = 0.013$, AUC: 0.983), with P4 × T^2 (AUC: 0.977) providing similar probability estimates (Fig. S1).

By determining cutoff values, we transformed continuous data into binary values for predicting whether an animal was pregnant or not, and we could then apply diagnostics statistics to determine the accuracy of each of the tests (Table 4). Diagnostic tests with an ROC AUC of greater than 0.90 can be considered excellent and would include an ultrasound exam (AUC: 0.992), observation of a calf

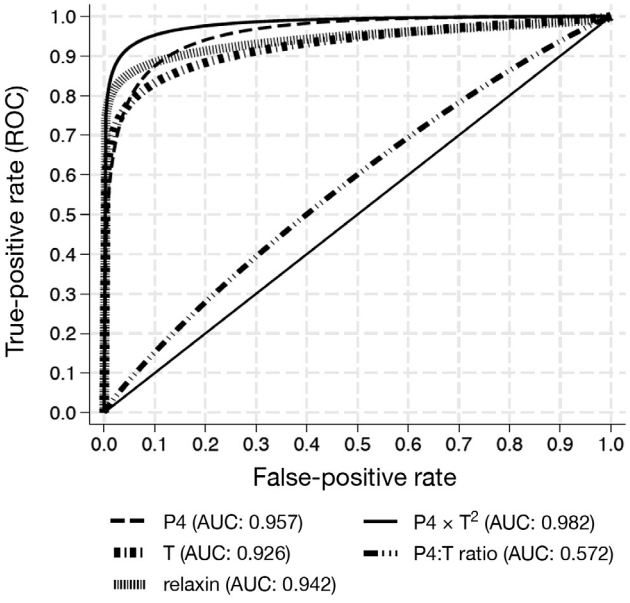

Fig. 2. Comparison of area under the curve (AUC) as calculated by receiver operator characteristic analysis (ROC) for serum progesterone (P4), testosterone (T), relaxin, and P4 × T^2 against pregnant or non-pregnant boto females using the gold standard. True positive rates (sensitivity) are plotted (ROC curves) versus false positive rates (1 − specificity). Solid black diagonal line: AUC of 0.50, representing a 50:50 chance of accurately predicting pregnancy. Hormones or hormone combinations with the highest AUC are considered to have the most accurate diagnostic potential. P4:T ratio appears only slightly better than chance and would not be of value. The gold standard represents a combination of positive pregnancy as determined by ultrasonography corrected for any false negative females that had been diagnosed as non-pregnant but were later observed with a calf of an age indicating that she had been pregnant during the preceding exam

(AUC: 0.934), and P4 × T^2 (AUC: 0.944), while tests with AUCs between 0.80 and 0.90 can be classified as good (Table 4) (Li & He 2018, Choe et al. 2023). Of all the hormone tests, P4 × T^2 had the highest positive predictive value (PPV) and negative predictive value (NPV) of 93.1 and 96.4%, respectively. This level of accuracy was only surpassed by ultrasound exam (PPV: 100%; NPV: 96.9%). For illustrative purposes, we then applied these predictions to the population they were developed from (with the knowledge that they may overestimate their accuracy compared to a novel population) to demonstrate how the results could be used to predict pregnancy and calving rates (Table S3). Large variations in the number of adult females examined each year (from 3 to 42) and possible sampling bias made year-to-year comparisons of reproductive rates inappropriate (Table S4). Therefore, to provide an estimate of the

Table 3. Receiver operator characteristic (ROC) area under the curve (AUC) values, comparisons, and 'cutoff' hormone concentrations (ng ml^{-1}) and the sensitivity (%), specificity (%), and correctly identified (% correct) values for detecting pregnancy compared to the 'gold standard' of pregnancy detection in boto (n = 126). Gold standard: ultrasound-diagnosed females with results corrected for false negatives if the female was later observed with a newborn calf (this occurred in only one case). P4: progesterone; T: testosterone; A4: androstenedione. BS: bootstrapped (1000 iterations); BC: bias-corrected; FPR: false positive rate

Analysis	Hormone				
	P4	T	A4	Relaxin	P × T^2
ROC AUC	0.9185	0.9337	0.5816	0.9252	0.9770
BS SE	0.0258	0.0226	0.0546	0.0272	0.0111
BS BC 95% CI	0.857–0.959	0.884–0.972	0.474–0.688	0.865–0.974	0.949–0.994
p-value[a]	–	0.66	<0.0001	0.75	0.0029
Cross-validated mean SD, BS, BC AUC	0.9248	0.9225	0.6137	0.9469	0.9721
Cross-validated SD	0.1086	0.1447	0.2233	0.0603	0.0693
Cross-validated BS BC 95% CI	0.815–0.949	0.839–0.967	0.422–0.656	0.832–0.960	0.878–0.985
pAUC (<10% FPR)	0.6393	0.8525	0.4590	0.7966	0.9344
BS SE	0.1615	0.0511	0.0801	0.0852	0.0412
p-value[b]	–	0.20	0.32	0.71	0.05
Cutoff	4.821	0.851	13.64	28.88	4.313
Sensitivity	91.8	88.5	45.9	93.2	93.4
Specificity	84.6	87.7	90.8	84.5	95.4
% Correct	88.1	88.1	69.05	88.9	94.4

[a]Probability value compared against the ROC AUC of P4
[b]Probablity value compared against the partial (pAUC <10% FPR) ROC pAUC of relaxin

Table 4. Diagnostic accuracy of survey observation, ultrasound exam, and serum markers for pregnancy in boto (n = 88) at a prevalence of 48% (CI: 39–58%) positive rate. ROCa: ROC area = (sensitivity + specificity) / 2; PPV (positive predictive value) = true positives / [true positives + false positives], NPV (negative predictive value) = true negatives / [true negatives + false negatives]

Test	Cutoff value	ROCa	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Observations						
Ultrasound exam	–	0.992	98.4 (91.2–100)	100 (94.5–100)	100 (94–100)	98.5 (91.8–100)
Calf observation	–	0.934	86.9 (75.8–94.2)	100 (94.5–100)	100 (93.3–100)	89 (79.5–95.1)
Single hormone						
Progesterone (P4)	4.821 ng ml^{-1}	0.874	90.2 (79.8–96.3)	84.6 (73.5–92.4)	84.6 (73.5–92.4)	90.2 (79.8–96.3)
Androstenedione	13.64 ng ml^{-1}	0.649	32.8 (21.3–46.0)	96.9 (89.3–99.6)	90.9 (70.8–98.9)	60.6 (70.8–98.9)
Testosterone (T)	0.853 ng ml^{-1}	0.873	86.9 (75.8–94.2)	87.7 (77.2–94.5)	86.9 (75.8–94.2)	87.7 (77.2–94.5)
Relaxin	28.88 ng ml^{-1}	0.844	93.4 (84.1–98.2)	75.4 (63.1–85.2)	78.1 (66.9–86.9)	92.5 (81.8–97.9)
Combinations						
P4:T ratio	3.06 nmol	0.538	49.2 (36.1–62.3)	58.5 (45.6–70.6)	52.6 (39.0–66.0)	55.1 (42.6–66.0)
P4 × T^2	4.313 ng ml^{-1}	0.944	98.4 (84.1–98.2)	95.4 (87.1–99.0)	95.0 (86.1–99.0)	93.9 (85.2–98.3)
Logit P4, T, and relaxin	0.502 ng ml^{-1}	0.898	95.1 (86.3–99.0)	84.6 (73.5–92.4)	85.3 (74.6–92.7)	94.8 (85.6–98.9)

population recruitment rate, we combined the data across all years and determined a pregnancy rate of 48% and a total calf loss rate of 13%.

3.4.3. Hormone reference ranges for female botos

Bootstrapped percentiles (2.5th, 50th, 97.5th) for P4, A4, T, Rlx, and P4 × T^2 within adult pregnant and non-pregnant females are shown in Table 5. Despite significant differences between pregnant and non-pregnant adult females, A4, T, and P4 in pregnant females overlapped with adult female concentrations; thus, lower values potentially observed during pregnancy could also occur during the luteal phase of non-pregnant females. Therefore, interpretation of sample concentrations from one of these hormones from a single time point may lead to an inaccurate or false positive diagnosis. As demonstrated by the ROC analysis, P4 × T^2 had the highest single-sample accuracy, with the least 95% CI overlap with non-pregnant adult females occurring during the last stage of preg-

Table 5. Reference intervals (2.5th, 50th and 97.5th percentiles) and the associated 95% bootstrap (1000 reps) confidence intervals (CI) for serum androstenedione (A4), testosterone (T), progesterone (P4), relaxin, and P4 × T^2 adult pregnant (n = 61) and non-pregnant (65) boto samples collected from 2003 through 2015. JF: juvenile female; AF: adult female (nonpregnant, nonlactating); PF: pregnant female. *analytes that differ ($p < 0.001$, except for A4 where $p = 0.015$) between AF and PF as determined by bootstrap *t*-test (1000 reps, $p < 0.05$) of 95% CI

Hormone (ng ml^{-1})	Female status	Percentile 2.5th (95% CI)	50th (95% CI)	97.5th (95% CI)
P4	AF	0.1 ± 0.03 (0.03 to 0.1)	1.2 ± 0.21 (0.8 to 1.5)	15.0 ± 2.4 (11.1 to 18.9)
	PF*	1.4 ± 0.7 (0.2 to 2.6)	9.72 ± 1.02 (8.05 to 11.40)	51.6 ± 14.4 (27.9 to 75.3)
A4	AF	3.9 ± 0.7 (2.7 to 5.0)	9.9 ± 0.5 (9.1 to 10.7)	18.3 ± 1.3 (16.1 to 20.5)
	PF*	0.8 ± 0.4 (0.1 to 1.6)	11.6 ± 1.6 (9.0 to 14.3)	32.0 ± 3.5 (26.3 to 37.8)
T	AF	0.1 ± 0.04 (0.1 ± 0.2)	0.6 ± 0.04 (0.6 to 0.7)	1.3 ± 0.1 (1.1 to 1.5)
	PF*	0.4 ± 0.1 (0.3 to 0.6)	3.8 ± 1.5 (1.3 to 6.2)	37.2 ± 5.6 (27.9 to 46.4)
Relaxin	AF	2.3 ± 3.2 (ND)	14.07 ± 0.08 (13.9 to 14.3)	193.3 ± 54.1 (104.4 to 282.3)
	PF*	14.1 ± 4.7 (6.3 to 21.9)	158.4 ± 27.4 (113.3 to 203.4)	1275.1 ± 373.0 (661.6 to 1888.7)
P4 × T^2	AF	0.01 ± 0.01 (ND)	0.4 ± 0.2 (0.14 to 0.63)	9.1 ± 1.7 (6.35 to 11.94)
	PF*	1.0 ± 0.8 (−0.35 to 2.3)	125.1 ± 161.7 (−140.9 to 391.1)	17447 ± 2985 (12538 to 22356)

nancy. These values, however, provide references for interpretation of the reproductive status of female botos and can help guide future analysis or sample collections.

3.4.4. Relationships between hormones

For males, as expected, Pearson pairwise correlation analysis detected strong relations ($r = 0.61$, $p < 0.0001$) between the androgens T and A4. For females, the strongest correlations existed between T and A4 ($r = 0.75$, $p < 0.0001$), and Rlx and T ($r = 0.74$, $p < 0.0001$). P4 was only moderately correlated with T ($r = 0.37$, $p = 0.0003$), Rlx ($r = 0.37$, $p = 0.0004$), and A4 ($r = 0.28$, $p = 0.006$). Finally, Rlx had a moderate correlation with A4 ($r = 0.48$, $p < 0.0001$).

4. DISCUSSION

This study is an extensive retrospective analysis of serum reproductive hormones in Amazon River dolphins (botos). Specifically, for males, we describe androgen concentration (T and A4) changes during maturation and estimate at what length these hormonal changes occur. For females, we describe multiple reproductive hormones, including P4, Rlx, T, and A4, in immature, adult non-pregnant, and pregnant botos. We also applied ROC techniques to evaluate the diagnostic potential of using single-sample analysis of each hormone, either individually or in combination, to predict whether or not an animal was pregnant.

Within males, we identified significant changes in both serum T and A4 concentrations based on the length and age class of the animals. Serum T has been well documented as an indicator of reproductive maturity and seasonal activity in cetaceans (Schroeder & Keller 1989, Robeck & Monfort 2006, Hao et al. 2007, O'Brien et al. 2008, 2017, Robeck et al. 2009). Within river dolphins, serum T was used to estimate that reproductive maturity occurs when male Yangtze River dolphins reached TLs of >138 cm (Hao et al. 2007). In the Amazon River dolphin, based on the distribution of serum T across TL, we estimated that animals were immature at a TL < 174 cm with serum T < 2.5 ng ml^{-1}, pubertal between 174 and 219 cm with serum T < 5 ng ml^{-1}, and adult at TBL > 219 cm with serum T > 4.9 ng ml^{-1}. Initial histologic analysis of testes found that sperm production was first noted for males from 190 to 198 cm in length (da Silva 1994, da Silva & Martin 2000, Martin & da Silva 2004, 2006). Additionally, previous work identified that the majority of scars indicative of male conspecific aggression, as typically noted in adult male boto, was observed in males >218 cm in length; results that are in line with our own (Martin & da Silva 2006).

Unlike serum T, A4 concentrations have only recently been characterized in a few species of male cetaceans, and evaluation of this hormone helps increase our understanding of T production (O'Brien et al. 2017). In both male and female botos, a strong correlation was observed between A4 and serum T,

supporting the premise that A4 is a precursor for T production, similar to what has been described for many domestic species and humans (Conley & Bird 1997). A4 has been documented to be formed within the adrenal cortex in higher primates and humans, whereas the testis are the primary source in the bull (Lindner & Rowson 1961) and boar (Schuler et al. 2014). Although the primary source of A4 within the male boto cannot be determined from these data, the pattern of correlation between serum T and A4 within male botos is like that observed in the bull and boar.

For females, the classification of sexually mature or/adult using P4 levels corroborates with previous studies for boto. In pregnant females, only A4 was not significantly different when compared to non-pregnant females. Nonetheless, a trend of increased A4 was observed as gestation progressed and is similar to killer whales and bottlenose dolphins, with the increase beginning by mid-gestation in the latter species (Robeck et al. 2017, Legacki et al. 2020, Steinman et al. 2021). Ultimately, the source of A4, whether a direct product of fetal gonad, luteal, or placental secretion is unknown (Legacki et al. 2020). However, evidence from retained corpora lutea in bottlenose dolphins suggests that androgens are not secreted from luteal tissue in any significant quantities and, therefore, a fetal and/or placental source is likely (Steinman et al. 2021).

Although androgens increased during gestation, our observation that the P4:T ratio was ineffective at differentiating pregnancy was unexpected. For the killer whale, this ratio was especially prominent and useful during early pregnancy for differentiating pregnant from non-pregnant females as well as identifying the stage of pregnancy (Robeck et al. 2017, Wasser et al. 2017). The lack of any significantly elevated P4:T ratio within the boto was due to lower than expected increases in P4 during early gestation combined with significantly increased T concentrations. These differences, when compared to other cetaceans, may represent a sampling artefact whereby a lack of early pregnancy samples (i.e. within 60 d of conception) were collected from the boto. In other cetaceans, P4 peaks and T slowly rises with placental development toward the end of this 60 d window. As we did not observe this trend, it may indicate that for the boto, serum T during the first trimester is derived primarily from luteal tissue. It is clear, however, that a diversity of steroid production during pregnancy, even within phylogenetically related species, is the rule rather than the exception (Conley et al. 2021), and only further research with the boto may help answer this question.

Surprisingly, concentrations of P4 appeared to be 2- to 3-fold lower for the boto than what has been observed in both the bottlenose dolphin (Robeck et al. 2021) and killer whale (Robeck et al. 2016). These may simply be due to different assay systems; however, P4 measurements across multiple platforms within cetaceans have been shown to be relatively consistent (O'Brien & Robeck 2012). Without further investigations, the reason for these differences in P4 concentration is unknown.

Like P4, Rlx was significantly elevated at each stage of pregnancy compared to non-pregnant adult females, but no differences between stages within pregnancy were observed. These observations are unlike those found in bottlenose dolphins and killer whales, whereby Rlx is significantly increased in each successive stage with a rapid 10-fold increase during the last month prior to parturition (Bergfelt et al. 2017, Robeck et al. 2018). These observed differences in the boto may simply be due to infrequent or missing samples toward the end of gestation when this surge is believed to occur, or it may be a true indication that the pattern of Rlx secretion in this species varies from other odontocetes. In support of the missing sample hypothesis for our data, we had one sample that was increased in Rlx concentrations (1710 ng ml^{-1}) well above the late-stage mean concentrations (289 ng ml^{-1}) in a female that was believed to have been close to parturition. This one sample lends support that like other cetaceans, Rlx does surge toward parturition in the boto.

Ultrasound diagnosis of pregnancy is considered the gold standard in multiple species, with FPRs during most of gestation approaching zero; however, during early pregnancy, detection is less accurate (Robeck et al. 2013, Abdullah et al. 2014, Ivančić et al. 2020). The day post ovulation when pregnancy can reliably be detected within each species is dependent on the rate of embryonic development, timing and development of placenta morphology, type of ultrasound equipment used, method (rectal versus transabdominal), and operator. For small cetaceans, between 44 and 60 d post ovulation appears to be the limit for consistent pregnancy detection using transabdominal ultrasonography (O'Brien & Robeck 2012, Robeck et al. 2013). For health assessments of wild animals, ultrasonography should be the primary tool. In seasonally breeding species, assessments can be scheduled during periods when pregnant animals would typically be near the end of the first stage or the beginning of the mid stage, so a fetus is easily detected (Wells et al. 2014, Lane et al. 2015). For species with diffuse or non-seasonal breeding like the

boto, a small percentage of animals will always be in early pregnancy, and false negatives are more likely. Ultimately, ultrasound assessment of ovaries (looking for the presence of a corpus luteum or follicles) and uterus (hypertrophy, atrophy, pregnancy) should be considered the most critical tool for reproductive exams during wild animal health assessments, and despite its high cost and a requirement of constant power source, every effort should be made to include it during all exams.

P4 had relatively low sensitivity compared to ultrasound for detecting pregnancy but was similar to the sensitivities of both T and Rlx. These hormones increase at different rates and stages during gestation, with a potential overlap with non-conceptive luteal phases during early pregnancy for P4 (O'Brien & Robeck 2012, Robeck et al. 2018). This overlap may, in part, explain the increased FPR (1 − specificity) for P4 (15%) which was almost 3 times that observed for P4 × T^2 (5%) and ~15 times that observed for ultrasound (0%). However, early embryonic loss between when maternal recognition of pregnancy (MRP) occurs (~15 d post conception in bottlenose dolphins) and prior to pregnancy detection by ultrasound may also be a contributing factor (O'Brien & Robeck 2012, Robeck et al. 2013, 2021). This period of embryo development may be susceptible to both external effects such as natural or anthropogenic environmental stressors (e.g. exposure to heat stress, endocrine disruptors, pathogens) that can lead to internal physiological disruption and is, therefore, an important period to understand within wild populations suffering reduced fecundity (Diskin & Morris 2008, Hansen 2014, Bellows & Short 2021, McGraw & Daigneault 2022). However, until an early pregnancy-specific factor can be identified, the normal or expected rate of loss during the period after MRP and prior to consistent ultrasound pregnancy detection remains speculative.

Besides P4, Rlx and T also had high FPRs, while A4 was relatively low. However, the inverse was true for their sensitivities. This is a reflection of the inverse relationship between sensitivity and specificity, whereby the FPR can be reduced with decreasing sensitivity based on the needs of the analysis or condition (pregnancy) being evaluated (Hart 2016). In the case of A4, very poor sensitivity (32%) resulted in a test with relatively low false positives (10%). In other words, concentrations of A4 below this relatively high concentration (13.64 ng ml^{-1} cutoff value) would have a 90% chance of not being pregnant. For population biologists, again, it is important to understand the rate of pregnancy loss and to therefore choose a test with the highest positive and negative predictive values. Test accuracy is important because pregnancy loss may be an indicator of exposure to natural or anthropogenic environmental stressors (Kight & Swaddle 2011). For example, a decrease in prey could cause pregnancy or calf loss due to inadequate maternal milk supply, or an increase in human activity (e.g. fishing) in critical nursery locations could influence calf survival (Kight & Swaddle 2011, Habeeb et al. 2018). Although heat stress is usually associated with land mammals, the exposure of boto to abnormally high water temperatures (Gonzalez-Socoloske 2023) should not be discounted or underestimated as to the potential impact that this environmental stressor may have on reproduction and calf survival.

We demonstrated the power of using the positive and negative prediction probabilities from each hormone to adjust the diagnostic results to improve the prediction of pregnancy and calving rates for this population of boto. This correction, based on probabilities, adjusts for the limitations of each of the hormone tests and was only able to be created due to the use of a highly accurate gold standard from which to develop these models. For individual animals, one would still have to use probabilities to estimate and differentiate between pregnant and non-pregnant animals. Negative and positive predictive values can change with the prevalence of the event that is being measured and can also be adjusted for clinical applications (Mandrekar 2010). For example, if one wants fewer false positives, the sensitivity can be adjusted by changing the cutoff point such that they only occur at an extremely low percentage and are usually due to the animal being in early pregnancy. However, by doing so, and depending on the accuracy of the test, this may cause an increase in false negatives.

The linear combination of T and P4 was extremely successful for classifying animals as pregnant or non-pregnant with high accuracy within the population. Although elevated Rlx can be considered pathognomonic for pregnancy (Bergfelt et al. 2017, Robeck et al. 2018), concentrations of Rlx in cetaceans do not reach these unequivocal levels until the end of the mid or late stage of pregnancy. As a result, we found relatively low specificity for Rlx, which translated into a low positive predictive value despite having a high NPV. In addition, limited availability of the Rlx RIA assay used in this study encouraged us to look for alternative hormone or hormone combinations, and P4 × T^2 was the most accurate. Indeed, the P4 × T^2 combination proved to be the most accurate hormone combination for PPV (95%) and NPV (93.9%), and as a

single-sample pregnancy test within the boto, it is close to the accuracy of ultrasound alone. Further evaluation of this combination for pregnancy prediction in other cetaceans is worth pursuing. The current results create a blueprint for developing reference ranges and diagnostic tests of other wild cetaceans, and using similar assay systems provides a means for estimating pregnancy rates in situations whereby ultrasound is not available.

Successful reproduction is often overlooked as an important variable for the long-term sustainability or recovery of long-lived, slow-growing mammalian populations like cetaceans (Manlik 2019). However, further analysis of survivability verses recruitment indicates that reproductive success can be paramount for population recovery and should be considered in population modeling or during the development of any wildlife conservation plans (Manlik 2019). The results of this research point to an improved capability to identify potential barriers toward successful recruitment within boto. By understanding at what stage reproductive failure occurs, further investigations can be devised to identify the etiology or etiologies of these failures and, when possible, implement mitigation factors. For example, if recruitment rates continue to decrease over time, increases in the frequency of observations (Booth et al. 2020) would improve resolution between primarily an increase in abortions verses poor neonatal survival. These 2 events would both result in reduced recruitment but would probably have divergent etiologies that would motivate different strategies, such as epidemiologic studies on potential environmental pathogens or toxins verses creating protection zones which may improve neonatal survival (Kellar et al. 2017, Cheney et al. 2019). As with most wildlife population strategies, increased surveillance will ultimately help with the decision process necessary to protect and preserve these and other endangered species.

In conclusion, our results demonstrate alternative ways to assess and measure reproductive maturation and pregnancy status in boto from a single blood sample. Hormonal analysis from a single time point using a combination of a suite of hormones provided increased accuracy for pregnancy diagnosis compared to traditional methods that have relied solely on a single hormone such as P4. As described within, ultrasonography was and will continue to be a valuable and critical tool for further developing these hormonal models for the application of diagnosing and predicting pregnancy outcomes in support of the conservation efforts for boto and other threatened species of cetaceans.

Acknowledgements. All procedures described herein were reviewed and approved by the National Institute of Amazonian Research (INPA) Animal Care and Use Committee. This project was approved by the following Brazilian government agencies: Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) and Sistema de Autorizacão em Biodiversidade (SISBio) (permit numbers 2344/9611-AC, 3552/9312-AC, 2002001.002344/96-11, 3462-1 to 13462-5, 13157-1 to 13157-4, and 50544-1 to 50544-3). Research was conducted under US National Marine Fisheries Service permit 16193 and Brazilian CITES export permit number 15BR018682/DF. This study was part of Projeto Boto, a cooperative agreement between the National Amazon Research Institute (INPA/MCTI) and the Mamirauá Sustainable Development Institute (MSDI-OS/MCTI). We thank Nivia do Carmo of LMA/INPA for help with sample collection. This is a SeaWorld Technical manuscript contribution number 2024-05.

LITERATURE CITED

Abdullah M, Mohanty T, Kumaresan A, Mohanty A, Madkar A, Baithalu R, Bhakat M (2014) Early pregnancy diagnosis in dairy cattle: economic importance and accuracy of ultrasonography. Adv Anim Vet Sci 2:464–467

Al-Sebai MAH, Kingsland CR, Diver M, Hipkin L, McFadyen IR (1995) The role of a single progesterone measurement in the diagnosis of early pregnancy failure and the prognosis of fetal viability. Br J Obstet Gynaecol 102:364–369

Bellows R, Short R (2021) Reproductive losses in the beef industry. In: Fields MJ, Sand RS (eds) Factors affecting calf crop. CRC Press, Boca Raton, FL, p 109–133

Bergfelt DR, Steinetz BG, Lasano S, West KL, Campbell M, Adams GP (2011) Relaxin and progesterone during pregnancy and the post-partum period in association with live and stillborn calves in bottlenose dolphins (*Tursiops truncatus*). Gen Comp Endocrinol 170:650–656

Bergfelt DR, Peter AT, Beg MA (2014) Relaxin: a hormonal aid to diagnose pregnancy status in wild mammalian species. Theriogenology 82:1187–1198

Bergfelt DR, Blum JL, Steinetz BG, Steinman KJ, O'Brien JK, Robeck TR (2017) Relaxin as a hormonal aid to evaluate pregnancy and pregnancy loss in bottlenose dolphins (*Tursiops truncatus*). Gen Comp Endocrinol 242:24–29

Booth CG, Sinclair RR, Harwood J (2020) Methods for monitoring for the population consequences of disturbance in marine mammals: a review. Front Mar Sci 7:115

Cheney BJ, Thompson PM, Cordes LS (2019) Increasing trends in fecundity and calf survival of bottlenose dolphins in a marine protected area. Sci Rep 9:1767

Choe H, Kobayashi N, Abe K, Hieda Y, Tezuka T, Inaba Y (2023) Evaluation of serum albumin and globulin in combination with C-reactive protein improves serum diagnostic accuracy for low-grade periprosthetic joint infection. J Arthroplasty 38:555–561

Clinical and Laboratory Standards Institute (2008) CLSI EP28. Defining, establishing and verifying reference intervals in the clinical laboratory, 3rd edn. CLSI document EP28-A3c. Clinical and Laboratory Standards Institute, Wayne, PA

Conley AJ, Bird I (1997) The role of cytochrome P450 17 α-hydroxylase and 3 β-hydroxysteroid dehydrogenase in

the integration of gonadal and adrenal steroidogenesis via the Δ5 and Δ4 pathways of steroidogenesis in mammals. Biol Reprod 56:789–799

Conley AJ, Loux SC, Legacki EL, Stoops MA and others (2021) The steroid metabolome of pregnancy, insights into the maintenance of pregnancy and evolution of reproductive traits. Mol Cell Endocrinol 528:111241

da Silva VMF (1994) Aspects of the biology of the amazonian dolphins *Inia* and *Sotalia fluviatilis*. PhD thesis, University of Cambridge

da Silva VMF, Martin AR (2000) A study of the boto, or Amazon river dolphin (*Inia geoffrensis*), in the Mamirauá Reserve, Brazil: operation and techniques. Occas Pap IUCN Species Survival Comm 23:121–130

da Silva VMF, Freitas CE, Dias RL, Martin AR (2018a) Both cetaceans in the Brazilian Amazon show sustained, profound population declines over two decades. PLOS ONE 13:e0191304

da Silva V, Trujillo F, Martin AR, Zerbini AN, Crespo E, Aliaga Rossel E, Reeves R (2018b) *Inia geoffrensis*. The IUCN Red List of Threatened Species 2018:e.T10831A 50358152. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T10831A50358152.en

da Silva VMF, Brum SM, de Mello DMD, de Souza Amaral R and others (2023) The Amazon River dolphin, *Inia geoffrensis*: What have we learned in the last two decades of research? Lat Am J Aquat Mamm 18:139–157

de Mello DMD, da Silva VMF (2019) Hematologic profile of Amazon river dolphins *Inia geoffrensis* and its variation during acute capture stress. PLOS ONE 14:e0226955

Diskin MG, Morris DG (2008) Embryonic and early foetal losses in cattle and other ruminants. Reprod Domest Anim 43:260–267

Dloniak SM, French JA, Place NJ, Weldele ML, Glickman SE, Holekamp KE (2004) Non-invasive monitoring of fecal androgens in spotted hyenas (*Crocuta crocuta*). Gen Comp Endocrinol 135:51–61

Faustini M, Battocchio M, Vigo D, Prandi A, Veronesi M, Comin A, Cairoli F (2007) Pregnancy diagnosis in dairy cows by whey progesterone analysis: an ROC approach. Theriogenology 67:1386–1392

Fricke PM (2002) Scanning the future — ultrasonography as a reproductive management tool for dairy cattle. J Dairy Sci 85:1918–1926

Gaona Calderón M, da Silva VMF, Barnett AA, Martin AR (2018) Estimating the length of dolphins using photographs where another animal of known or estimated length is in close proximity. Mar Mamm Sci 34: 1111–1118

Gonzalez-Socoloske D (2023) An unexpected event related to the aquatic mammals of Latin America. Lat Am J Aquat Mamm 18:167–168

Graham L, Schwarzenberger F, Möstl E, Galama W, Savage A (2001) A versatile enzyme immunoassay for the determination of progestogens in feces and serum. Zoo Biol 20: 227–236

Graham LH, Bolling J, Miller G, Pratt-Hawkes N, Joseph S (2002) Enzyme-immunoassay for the measurement of luteinizing hormone in the serum of African elephants (*Loxodonta africana*). Zoo Biol 21:403–408

Habeeb AA, Gad AE, Atta MA (2018) Temperature–humidity indices as indicators to heat stress of climatic conditions with relation to production and reproduction of farm animals. Int J Biotechnol Recent Adv 1:35–50

Hansen PJ (2014) Early embryonic loss due to heat stress. In: Hopper RM (ed) Bovine reproduction. John Wiley & Sons, Ames, IA, p 580–588

Hao YJ, Chen DQ, Zhao QZ, Wang D (2007) Serum concentrations of gonadotropins and steroid hormones of *Neophocaena phocaenoides asiaeorientalis* in middle and lower regions of the Yangtze River. Theriogenology 67: 673–680

Hart PD (2016) Receiver operating characteristic (ROC) curve analysis: a tutorial using body mass index (BMI) as a measure of obesity. J Phys Act Res 1:5–8

Hildebrandt TB, Brown JL, Hermes R, Goritz F (2003) Ultrasound for analysis of reproductive function in wildlife species. In: Hildebrandt TB, Brown JL, Hermes R, Göritz F (eds) Reproductive science and integrated conservation. Cambridge University Press, Cambridge, p 166–182

Ivančić M, Gomez FM, Musser WB, Barratclough A and others (2020) Ultrasonographic findings associated with normal pregnancy and fetal well-being in the bottlenose dolphin (*Tursiops truncatus*). Vet Radiol Ultrasound 61: 215–226

Janes H, Longton G, Pepe MS (2009) Accommodating covariates in receiver operating characteristic analysis. Stata J 9:17–39

Kellar NM, Speakman TR, Smith CR, Lane SM and others (2017) Low reproductive success rates of common bottlenose dolphins *Tursiops truncatus* in the northern Gulf of Mexico following the *Deepwater Horizon* disaster (2010–2015). Endang Species Res 33:143–158

Kight CR, Swaddle JP (2011) How and why environmental noise impacts animals: an integrative, mechanistic review. Ecol Lett 14:1052–1061

Lacave G, Eggermont M, Verslycke T, Brook F, Salbany A, Roque L, Kinoshita R (2004) Prediction from ultrasonographic measurements of the expected delivery date in two species of bottlenosed dolphin (*Tursiops truncatus* and *Tursiops aduncus*). Vet Rec 154:228–233

Lane SM, Smith CR, Mitchell J, Balmer BC and others (2015) Reproductive outcome and survival of common bottlenose dolphins sampled in Barataria Bay, Louisiana, USA, following the *Deepwater Horizon* oil spill. Proc R Soc B 282:20151944

Legacki EL, Robeck TR, Steinman KJ, Conley AJ (2020) Comparative analysis of steroids in cyclic and pregnant killer whales, beluga whales and bottlenose dolphins by liquid chromatography tandem mass spectrometry. Gen Comp Endocrinol 285:113273

Li F, He H (2018) Assessing the accuracy of diagnostic tests. Shanghai Arch Psychiatry 30:207

Lindner HR, Rowson LEA (1961) Androgens and related compounds in the spermatic vein blood of domestic animals. III. The effect of unilateral orchidectomy on the rate of androgen secretion by the remaining testis in the immature calf. J Endocrinol 23:167–170

Linnet K (2000) Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. Clin Chem 46:867–869

Luque-Fernandez MA, Redondo-Sánchez D, Maringe C (2019) cvauroc: command to compute cross-validated area under the curve for ROC analysis after predictive modeling for binary outcomes. Stata J 19:615–625

Mandrekar JN (2010) Simple statistical measures for diagnostic accuracy assessment. J Thorac Oncol 5:763–764

Manlik O (2019) The importance of reproduction for the conservation of slow-growing animal populations. In: Comiz-

zoli P, Brown JL, Holt WV (eds) Reproductive sciences in animal conservation. Springer, Cham, p 13–39
Martin AR, da Silva VMF (2004) River dolphins and flooded forest: seasonal habitat use and sexual segregation of botos (*Inia geoffrensis*) in an extreme cetacean environment. J Zool 263:295–305
Martin AR, da Silva VMF (2006) Sexual dimorphism and body scarring in the boto (Amazon river dolphin) *Inia geoffrensis*. Mar Mamm Sci 22:25–33
Martin AR, da Silva VMF (2018) Reproductive parameters of the Amazon river dolphin or boto, *Inia geoffrensis* (Cetacea: Iniidae); an evolutionary outlier bucks no trends. Biol J Linn Soc 123:666–676
Martin AR, da Silva VMF (2022) Amazon river dolphins *Inia geoffrensis* are on the path to extinction in the heart of their range. Oryx 56:587–591
Martin AR, da Silva VMF, Salmon D (2004) Riverine habitat preferences of botos (*Inia geoffrensis*) and tucuxis (*Sotalia fluviatilis*) in the central Amazon. Mar Mamm Sci 20: 189–200
Mazzoil M, McCulloch SD, Defran RH (2005) Observations on the site fidelity of bottlenose dolphins (*Tursiops truncatus*) in the Indian River Lagoon, Florida. Fla Sci 68: 217–226
McGraw MS, Daigneault BW (2022) Environment to embryo: intersections of contaminant exposure and preimplantation embryo development in agricultural animals. Biol Reprod 107:869–880
Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM (2012) Quantifying the added value of a diagnostic test or marker. Clin Chem 58:1408–1417
Munro CJ, Lasley BL (1988) Non-radiometric methods for immunoassay of steroid hormones. Prog Clin Biol Res 285:289–329
Nabi G, Robeck TR, Hao Y, Wang D (2019) Hematologic and biochemical reference interval development and the effect of age, sex, season, and location on hematologic analyte concentrations in critically endangered Yangtze finless porpoise (*Neophocaena asiaeorientalis* ssp. *asiaeorientalis*). Front Physiol 10:792
O'Brien JK, Robeck TR (2012) The relationship of maternal characteristics and circulating progesterone concentrations with reproductive outcome in the bottlenose dolphin (*Tursiops truncatus*) after artificial insemination, with and without ovulation induction, and natural breeding. Theriogenology 78:469–482
O'Brien JK, Steinman KJ, Schmitt T, Robeck TR (2008) Semen collection, characterisation and artificial insemination in the beluga (*Delphinapterus leucas*) using liquid-stored spermatozoa. Reprod Fertil Dev 20:770–783
O'Brien JK, Steinman KJ, Fetter GA, Robeck TR (2017) Androgen and glucocorticoid production in the male killer whale (*Orcinus orca*): influence of age, maturity, and environmental factors. Andrology 5:180–190
Pepe MS, Longton G, Janes H (2009) Estimation and comparison of receiver operating characteristic curves. Stata J 9: 1–16
Robeck TR, Monfort SL (2006) Characterization of male killer whale (*Orcinus orca*) sexual maturation and reproductive seasonality. Theriogenology 66:242–250
Robeck TR, Steinman KJ, Greenwell M, Ramirez K and others (2009) Seasonality, estrous cycle characterization, estrus synchronization, semen cryopreservation, and artificial insemination in the Pacific white-sided dolphin (*Lagenorhynchus obliquidens*). Reproduction 138:391–405
Robeck TR, Montano G, Steinman K, Smolensky P, Sweeney J, Osborn S, O'Brien J (2013) Development and evaluation of deep intra-uterine artificial insemination using cryopreserved sexed spermatozoa in bottlenose dolphins (*Tursiops truncatus*). Anim Reprod Sci 139:168–181
Robeck TR, Schmitt TL, Osborn S (2015) Development of predictive models for determining fetal age-at-length in belugas (*Delphinapterus leucas*) and their application toward *in situ* and *ex situ* population management. Mar Mamm Sci 31:591–611
Robeck TR, Steinman KJ, O'Brien JK (2016) Characterization and longitudinal monitoring of serum progestagens and estrogens during normal pregnancy in the killer whale (*Orcinus orca*). Gen Comp Endocrinol 236:83–97
Robeck TR, Steinman KJ, O'Brien JK (2017) Characterization and longitudinal monitoring of serum androgens and glucocorticoids during normal pregnancy in the killer whale (*Orcinus orca*). Gen Comp Endocrinol 247:116–129
Robeck TR, Blum JL, Steinman KJ, Ratner JR, Bergfelt DR, O'Brien JK (2018) Longitudinal profiles of relaxin and progestagens during pregnancy, pregnancy loss and false pregnancy in the killer whale (*Orcinus orca*). Gen Comp Endocrinol 267:98–108
Robeck TR, Amaral RS, da Silva VMF, Martin AR, Montano GA, Brown JL (2019) Thyroid hormone concentrations associated with age, sex, reproductive status and apparent reproductive failure in the Amazon river dolphin (*Inia geoffrensis*). Conserv Physiol 7:coz041
Robeck TR, Steinman KJ, Parry CB, Gomez FM, Jensen ED (2021) Comparisons of serum progesterone and progestagen concentrations in normal and abnormal bottlenose dolphin (*Tursiops truncatus*) pregnancies. Front Mar Sci 8:630563
Saeed S, Alireza B, Mohamed E, Ahmed N (2015) Evidence based emergency medicine part 2: positive and negative predictive values of diagnostic tests. Emergency 3: 87–88
Schroeder JP, Keller KV (1989) Seasonality of serum testosterone levels and sperm density in *Tursiops truncatus*. J Exp Zool 249:316–321
Schuler G, Dezhkam Y, Bingsohn L, Hoffmann B and others (2014) Free and sulfated steroids secretion in postpubertal boars (*Sus scrofa domestica*). Reproduction 148: 303–314
Seed PTT, Tobias A (2001) Summary statistics for diagnostic tests. Stata Tech Bull 59:9–12
Speakman T (2006) Temporal and spatial aspects of bottlenose dolphin occurrence in coastal and estuarine waters near Charleston, South Carolina. NOAA Tech Memo NOS-NCCOS-37
Steinman KJ, Robeck TR, O'Brien JK (2016) Characterization of estrogens, testosterone, and cortisol in normal bottlenose dolphin (*Tursiops truncatus*) pregnancy. Gen Comp Endocrinol 226:102–112
Steinman K, Montano G, Robeck T (2021) Characterization of circulating androgens, cortisol and estrogens during normal, abnormal and false pregnancy in bottlenose dolphins (*Tursiops truncatus*) under managed care. Front Mar Sci 8:737926
Trego ML, Kellar NM, Danil K (2013) Validation of blubber progesterone concentrations for pregnancy determination in three dolphin species and a porpoise. PLOS ONE 8:e69709
Urian KW, Wells RS (1996) Bottlenose dolphin photo-identification workshop, March 21–22, 1996, Charleston,

South Carolina: final report to the National Marine Fisheries Service, Charleston Laboratory. NOAA Tech Memo NMFS-SEFSC-393

Urian K, Gorgone A, Read A, Balmer B and others (2015) Recommendations for photo-identification methods used in capture–recapture models with cetaceans. Mar Mamm Sci 31:298–321

Wasser SK, Lundin JI, Ayres K, Seely E and others (2017) Population growth is limited by nutritional impacts on pregnancy success in endangered southern resident killer whales (*Orcinus orca*). PLOS ONE 12:e0179824

Wells RS, Smith CR, Sweeney JC, Townsend FI and others (2014) Fetal survival of common bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida. Aquat Mamm 40:252

West BT, Welch KB, Galecki AT (2014) Linear mixed models: a practical guide using statistical software. CRC Press, Boca Raton, FL

Editorial responsibility: Brendan Godley, University of Exeter, Cornwall Campus, UK
Reviewed by: 2 anonymous referees

Submitted: May 7, 2024
Accepted: July 9, 2024
Proofs received from author(s): August 4, 2024