A COMPARATIVE EVALUATION OF FOUR FIELD METHODS FOR SEXING WEDGE-TAILED SHEARWATERS *PUFFINUS PACIFICUS*

STEPHEN L. TOTTERMAN

179 Reedy Creek Road, Empire Vale, NSW 2478, Australia (stephen@totterman.net.au)

Received 14 July 2014, accepted 27 November 2014

SUMMARY

TOTTERMAN, S.L. 2015. A comparative evaluation of four field methods for sexing Wedge-tailed Shearwaters *Puffinus pacificus*. *Marine Ornithology* 43: 83–93.

Identifying female and male birds can be very helpful in field studies. However, sexual differences in size and plumage are subtle in most petrels. Four field methods were compared for sexing breeding Wedge-tailed Shearwaters *Puffinus pacificus* on Muttonbird Island, New South Wales, Australia: cloaca inspection, biometrics, acoustics and playback-response. Accuracy was evaluated against molecular tests. A biometric discriminant function combining bill depth and total head length sexed 81% of birds (79 of 98) correctly. Males averaged 3% larger than females, with overlapping size ranges. Sexual differences in cloacal size were not always obvious because female cloacae gradually relapse after laying and males struggling in the hand can present extruded cloacae. Cloacal sexing was 86% correct (93 of 108 birds). Within-pair comparisons of biometrics and cloacal size increased sex classification accuracy for twice the effort (two birds evaluated rather than one). An acoustic discriminant function combining fundamental frequency and note length from burrow call recordings sexed 97% of birds (102 of 105) correctly. A novel playback-response test was efficient and sexed 94% of birds (47 of 50) correctly.

Key words: acoustic, biometric, cloaca inspection, discriminant function, playback, sexing, sexual size dimorphism

INTRODUCTION

Many petrels exhibit ecological and behavioural differences between the sexes (Warham 1990). However, field identification of sexes can be difficult because differences in female and male size and plumage are subtle (e.g. Bull *et al.* 2005). Popular field sexing methods used have been cloaca inspection (Serventy 1956) and biometrics (e.g. Thalmann *et al*. 2007, Einoder *et al*. 2008, Carey 2011). Vocalisations have been used less frequently because sexual differences in calls are poorly documented for most petrels.

The petrel egg is very large relative to body size, and breeding females show an enlarged cloaca near laying. Serventy (1956) suggested that cloaca inspection can distinguish breeding females during the period from about three weeks before to four weeks after laying. O'Dwyer *et al.* (2006) claimed that cloacal sexing can be applied to non-breeding petrels as well as breeders, which is contradicted in their reference (Bartle 1968), which stated that "non-breeding birds cannot reliably be sexed by this method." For example, Copestake *et al.* (1988) measured cloacal size to separate breeding females from others in samples of Wilson's Storm-Petrels *Oceanites oceanicus* caught in mist nets. Although cloacal sexing is widely used for petrels and assumed to be reliable, there have been few published, quantitative evaluations of the method (Boersma & Davies 1987, O'Dwyer *et al.* 2006).

Males are on average slightly larger than females in most petrels (Warham 1990). Male-biased sexual size dimorphism (SSD) in shearwaters from the genus *Puffinus* is <5% (i.e. males 5% larger than females), except for bill depth measurements in some species (Bull *et al.* 2005). Multivariate discriminant functions can improve biometric sexing accuracy when there is overlap in size ranges (Dechaume-Moncharmont *et al.* 2011). Alternatively, simple within-breeding-pair comparisons of size can classify sexes with accuracy comparable to biometric discriminant functions (e.g. Genovart *et al.* 2003, Bourgeois *et al.* 2007, Carey 2011).

Most petrels are nocturnal at their breeding colonies and primarily use vocalisations to communicate in the dark (Shallenberger 1973, Bretagnolle 1996). Nine shearwaters from the genus *Puffinus* have a single major call, used for both courtship and in territorial contacts and given mostly from the ground (Bretagnolle 1996). Major calls of males are typically clearer, higher pitched and have longer note durations than those of females (Brooke 1978, 1988; James & Robertson 1985, Bretagnolle *et al.* 2000, Bourgeois *et al.* 2007). For some petrels, differences in female and male calls can be recognised by human listeners (e.g. Brooke 1978, James & Robertson 1985, Bretagnolle *et al.* 2000, Totterman 2012), but for others quantitative acoustic analysis is recommended (e.g. Brooke 1988, Taoka *et al.* 1989b). Playback-response experiments have been used to demonstrate vocal sex recognition in petrels. Incubating shearwaters typically respond only to same-sex calls, but may also respond to calls of their mates (e.g. Brooke 1978, 1988; Cure *et al.* 2009).

Totterman (2014) described sexual differences in Wedge-tailed Shearwater burrow calls and demonstrated vocal sex recognition. Male calls averaged significantly higher in fundamental frequencies and longer in note lengths than those of females, with overlapping ranges. Incubating females nearly exclusively responded to female playbacks whereas males responded to both female and male calls. This study investigated acoustic discriminant function sexing and a playback-response method for testing burrow occupancy and sex in Wedge-tailed Shearwaters. Results are compared with cloaca inspection and biometric methods.

STUDY AREA AND METHODS

Breeding Wedge-tailed Shearwaters were studied from 9 December 2012 to 6 January 2013, during the incubation phase, on Muttonbird Island (13°48′S, 167°29′E), New South Wales, Australia. All field work, recording and playback-response experiments were performed at night. Each study bird was uniquely identified by its burrow and two non-permanent markings (one or two spots of white paint on the crown and left or right outer tail feather clipped). I could not use uniquely numbered leg bands because Australian banding regulations are quite onerous. Further details about the study site and field methods are described in Totterman (2014).

Molecular sexing

For molecular tests, a few breast feathers were sampled from each study bird and kept in separate, sealed plastic bags. These bags were stored together in a sealed container with silica gel desiccant. Sex was determined by polymerase chain reaction (PCR) methods sensitive to markers on the *CHD* gene. The primary laboratory was located at the University of Queensland, Australia, where the primer pair P2/P3 was used, followed by restriction enzyme digestion with *HaeIII* (Norris-Caneda & Elliott 1998). A duplicate verification sub-sample (Robertson & Gemmel 2006) was analysed at Gribbles Veterinary Australia, Clayton, Australia, using the primer pair P2/P8 (Griffiths *et al.*1998). Both laboratories analysed PCR products by capillary electrophoresis. Positive controls (one known-sex pair) were provided and initially used to evaluate primers and optimise sample preparation, PCR and electrophoresis conditions. Positive controls were subsequently used for orientation and quality control. These known-sex positive controls were breeding pairs with large differences in cloacal size and contrasting responses to playback-response tests. If these positive controls were incorrectly sexed, which did not occur, this would have been immediately detected as frequent unexpected banding patterns from molecular tests.

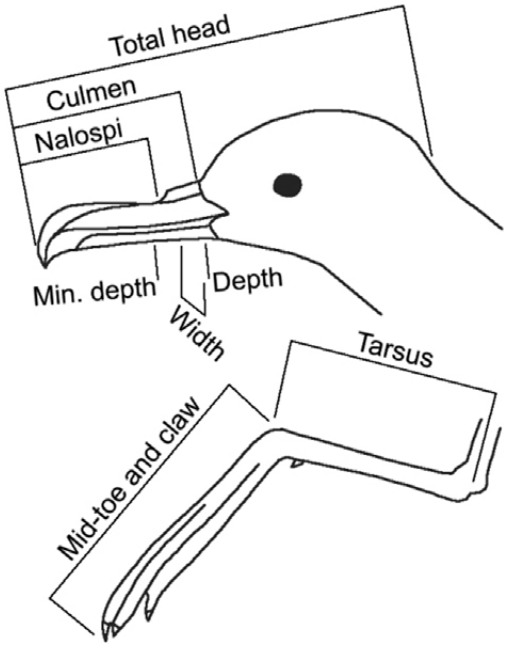

Fig. 1. Head, bill and leg measurements taken from each Wedge-tailed Shearwater. Additional biometrics not illustrated are wing length, tail length and mass.

Cloaca inspection

Cloaca inspection was performed on each bird before measuring biometrics. First, an overall assessment of size, shape and prominence of lips around the cloacal opening was made, and cloacae were subsequently categorized into seven size classes: extra-small, small, small–medium, medium, medium–large, large and extra-large. Second, the width of the cloacal opening was measured transversely across the body with a vernier caliper (rounded to 0.5 mm; O'Dwyer *et al.* 2006). More precise cloacal width measurements were not attempted because of birds struggling in the hand, cloacal plasticity and obstructing feathers (I was working alone). Cloaca inspections were repeated after five to 24 days on the first 13 birds processed, when I had become familiar with handling the birds and variation in cloacal size and shape.

Biometrics

Biometrics measured with a vernier caliper (resolution 0.1 mm) were total head length, exposed culmen length (hereafter "culmen"), culmen length to the anterior edge of the nostrils (hereafter "nalospi"), bill depth at the base of the exposed culmen (hereafter "bill depth"), minimum bill depth, bill width (at the base of the exposed culmen), tarsometatarsus length (hereafter "tarsus") and mid-toe and claw length (Fig. 1). Tail length was measured by gently pushing the outer jaw of the caliper between the central retrices. A modified "head caliper," with a 10 × 14 mm flat pad on the inner jaw, was used to measure total head length. Wing length (maximum flattened chord) was measured with a steel wing ruler (resolution 1 mm). Mass was measured with a 0–500 g spring balance (resolution 5 g). Weighing bags were measured with a 0–50 g spring balance (resolution 0.5 g) immediately after removing each bird.

Acoustics

Playbacks of same-sex calls to incubating Wedge-tailed Shearwaters would often result in vigorous responses ("defence calls" in Shallenberger 1973). These major calls consist of repeated two-note units ("syllables"; following the terminology of Thompson *et al.* 1994) that sound like "*ooh-ah*" or "*kooh-ah*" (Fig. 2). Exhalant

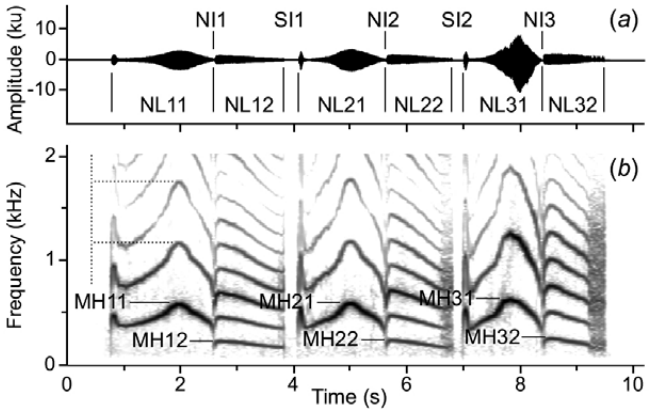

Fig. 2. Waveform (a) and spectrogram (b) of the first three syllables of an adult male Wedge-tailed Shearwater burrow call. Call measurements annotated are: note lengths (NL11, 12, 21, 22, 31, 32), note intervals (NI1, 2, 3), syllable intervals (SI1, 2) and peak fundamental frequencies (mean harmonic intervals MH11, 12, 21, 22, 31, 32).

first notes are louder, higher-pitched and longer than inhalant second notes. Successive syllables increase in volume and pitch to a peak and then decline towards the end of a call. Late syllables and especially inhalant notes can be noisy (Shallenberger 1973).

Burrow calls were recorded with a Sony PCM-M10 recorder (16 bit, 22.05 kHz uncompressed digital audio) and Sennheiser ME 64 directional microphone. Acoustic measurements were made in Raven Pro version 1.3 (Cornell Laboratory of Ornithology, Ithaca). Temporal-domain measurements from call waveforms were note lengths, note intervals and syllable intervals (Fig. 2a). Frequency-domain measurements from spectrograms were peak mean harmonic intervals of each note (Fig. 2b). Mean harmonic interval estimates fundamental frequency. The spectrogram 3 decibel filter bandwidth was 16 Hz. Two syntactic features were measured: the total number of syllables per call and the syllable number at the crescendo peak. Analyses focused on syllables one to three because 1) birds often responded immediately during playback, 2) a few birds gave mostly three-syllable calls, 3) the most frequent number of syllables per call was three or four, and 4) the crescendo most frequently peaked in syllable three (Totterman 2014).

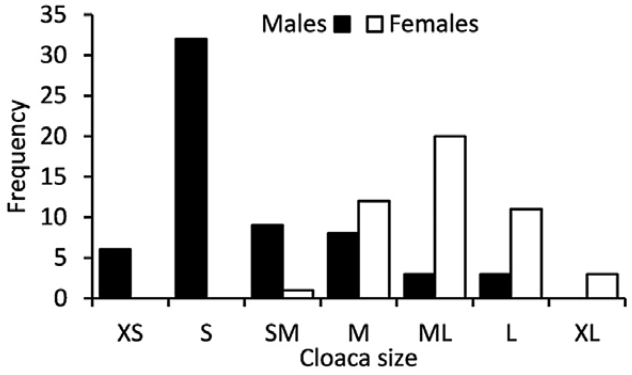

Fig. 3. Cloacal size distributions for breeding Wedge-tailed Shearwaters from visual assessments: XS = extra-small, S = small, SM = small–medium, M = medium, ML = medium–large, L = large and XL = extra-large. Sizes M to XL were classified as females and XS to SM as males. See results for cloacal width measurements.

Playback-response tests

I played Wedge-tailed Shearwater calls through a small battery-powered speaker. The speaker was placed on the ground, facing the study burrow entrance, and the sound volume was approximately matched by ear to the level of natural calls.

At each active study burrow I played back two calls from two different males, followed by two calls from two different females. I waited for any responses to complete between playing back the next call in sequence. Subjects that responded exclusively to female calls or not at all were classified as females. Subjects that responded to one or both male calls were classified as males (Totterman 2014). The burrow occupant was then captured and identified by its burrow and markings.

Twenty three female and 23 male recordings, each edited to 10 to 16 s duration, were rotated between tests. The sex of the birds in these recordings was established by within-pair comparisons of cloacal size and subsequently verified with molecular results. Only unfamiliar ("stranger") calls were played back, because petrels are expected to respond differently towards neighbours (e.g. Mackin 2005, Totterman 2014), mates (e.g. Brooke 1978, Cure *et al.* 2009) and to playbacks of their own calls (Shallenberger 1973).

TABLE 1
Wedge-tailed Shearwater cloacal sexing accuracy from visual assessments of size and from width measurements

	Statistic	Females (n = 47)	Males (n = 61)	Pairs (n = 39)	(F–M) difference	(Pairs–F) difference[a]	(Pairs–M) difference[a]
Cloacal size	Prop. (%)	98	77	92	21	−5	13
	95% CI (%)	89–100	65–86	80–97	9–32	−18 to 7	0.8–26
	Test[b]				$\chi^2_1 = 9.5$	X = 3, n = 4	X = 5, n = 5
	P				0.002	0.38	0.03
Cloacal width	Prop. (%)	89	84	90	6	0.0	3
	95% CI (%)	77–95	72–91	76–96	−7 to 19	−11 to 11	−9 to 15
	Test[a]				$\chi^2_1 = 0.73$	X = 2, n = 4	X = 2, n = 3
	P				0.39	0.81	0.63
Size – width difference	Prop. (%)	9	−7	3			
	95% CI (%)	−1.3 to 20	−15 to 0.8	−7 to 13			
	Test[b]	X = 4, n = 4	X = 4, n = 4	X = 1, n = 1			
	P	0.06	0.06	0.50			

[a] Female and male proportions compared in (Pairs–F) and (Pairs–M) refer to the breeding pairs sample (n = 39) and are −3.6% to +0.4% different to the female and male proportions in the total sample.

[b] "N – 1" *chi*-squared tests for independent proportions and McNemar mid-*P* tests for paired proportions.

Statistical analyses

Statistical analyses were performed with R version 3.0.1 (R Development Core Team 2013). Biometrics were screened for outliers using dot plots. A few unusual observations were deleted (e.g. suspected curvature of the toe). Further checks included boxplots by sex to evaluate homogeneity of variance, quantile-quantile plots to evaluate normality and scatterplots to evaluate collinearity (Zuur *et al.* 2010). SSD was evaluated using two-sample *t*-tests. The acoustic measurements dataset was checked, and sexual differences in burrow calls evaluated in Totterman (2014).

Linear discriminant functions were computed with the lda function from the MASS R library version 7.3-26 (Venables & Ripley 2002). Selection of predictor variables was stepwise forward with manual selection. Classification accuracy was evaluated by leave-one-out cross-validation (Lachenbruch & Mickey 1968). Additional predictors were included only if they increased accuracy by at least 2%. Multicollinearity can confound discriminant analyses (Dechaume-Moncharmont *et al.* 2011), and combinations of strongly correlated variables ($r \geq 0.7$) were therefore not considered.

Comparisons between independent proportions (i.e. proportions of correct sex classifications) were made using "N − 1" *chi*-squared tests (Pearson 1947; evaluated by Campbell 2007). Paired results were compared using McNemar mid-P tests (evaluated by Fagerland *et al.* 2013). There were no *a priori* expectations that any field sexing method was more accurate than another and two-sided *P* values are reported throughout this paper.

Precision for single proportions was estimated by Wilson score intervals (Wilson 1927; evaluated by Newcombe 1998a). Precision for differences between independent proportions was estimated by Newcombe's (1998b; Method 10) combined Wilson score intervals. These were calculated with the prop.test R function without continuity correction. Precision for differences between paired proportions was estimated by Newcombe's (1998c; Method 10) combined Wilson score intervals for paired data. This was calculated with the wilson.phi function from the diffdepprop R library version 0.1-9 (Wenzel & Zapf 2013). These score intervals are approximate and were selected for good mean coverage probability.

Assortative mating for biometrics and acoustics was evaluated using Spearman rank correlations. An overall body size index was estimated from the first principal component for biometrics (Rising & Somers 1989). Principal components were computed using singular value decomposition of the centred and scaled data matrix with the princomp R function. Additionally, observed within-pair

TABLE 2
Wedge-tailed Shearwater biometrics and sexual size dimorphism (SSD) calculated as the difference between male and female means expressed as percent of the female mean

						Mean difference				
Biometric	**Sex**	**n**	**Mean**	**SD**	**Range**	**(M–F)**	**95% CI**	t[a]	**P**	**SSD (%)**
Wing (mm)	M	51	288	6	278–299	3	0.7–6	2.6	0.012	1
	F	47	285	6	274–299					
Tail (mm)	M	51	131	4	123–140	−1	−3 to 0.5	−1.4	0.18	−1
	F	46	132	4	123–142					
Head (mm)	M	51	85.6	1.7	81.5–88.7	2.2	1.6–2.8	7.0	<0.001	3
	F	47	83.4	1.4	79.4–86.6					
Culmen (mm)	M	51	38.0	1.3	35.4–40.8	1.3	0.8–1.8	5.6	<0.001	4
	F	47	36.7	0.9	34.3–39.2					
Nalospi (mm)	M	51	28.9	1.1	26.4–31.5	1.1	0.7–1.5	5.4	<0.001	4
	F	47	27.8	0.8	26.1–30.1					
Bill depth (mm)	M	51	13.2	0.5	12.4–14.2	0.6	0.5–0.8	7.0	<0.001	5
	F	47	12.5	0.4	11.9–13.6					
Min. bill depth (mm)	M	51	9.2	0.4	8.4–10.3	0.5	0.4–0.7	7.0	<0.001	6
	F	47	8.6	0.3	7.9–9.4					
Bill width (mm)	M	50	12.9	0.4	12.0–14.0	0.4	0.2–0.5	4.0	<0.001	3
	F	46	12.5	0.4	11.7–13.4					
Tarsus (mm)	M	51	49.2	1.3	45.9–51.6	0.7	0.2–1.2	2.7	0.009	1
	F	47	48.5	1.3	46.0–51.6					
Mid-toe & claw (mm)	M	50	60.1	1.9	55.4–64.6	1.0	0.2–1.7	2.7	0.009	2
	F	46	59.2	1.5	55.2–62.2					
Mass (g)	M	51	375	32	318–437	−6	−18 to 7	−0.93	0.36	−2
	F	44	381	29	302–447					

[a] Degrees of freedom for the two-sample *t*-tests equal the number of females plus number of males minus two.

comparisons were compared with those from random female-male pairs using non-parametric bootstrapping with 999 resamples.

RESULTS

Molecular sexing

A total of 145 molecular sex tests were performed for 108 shearwaters (47 females, 61 males). Results for all 36 birds (nine females, 27 males) in the verification subsample verified correctly at the secondary laboratory. Results for all 37 complete breeding pairs were consistent with heterosexual social monogamy expected for petrels (Warham 1990). There were two PCR amplification failures, which fortunately occurred in breeding pairs in which both birds were sampled. The complementary sex was assumed for each of these missing results. Subsequently, accuracy of field sexing methods was evaluated against molecular results.

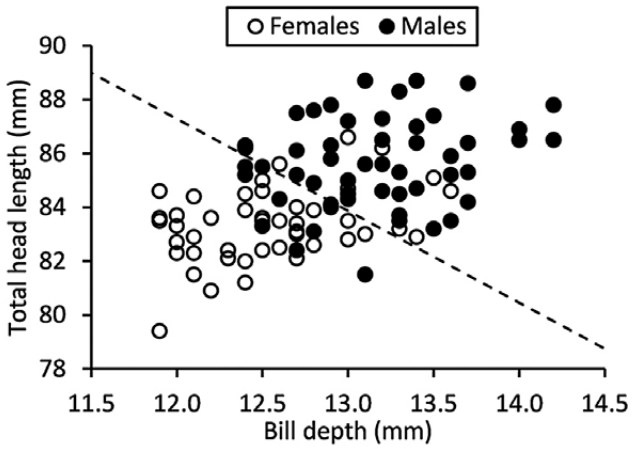

Fig. 4. Total head length and bill depth for Wedge-tailed Shearwaters (n = 47 females, 51 males). The dashed line indicates equal female/male probability from the biometric discriminant function (DS_1 = 0). Individuals plotting above the line ($DS_1 > 0$) are classified as males and individuals below ($DS_1 < 0$) as females.

Cloaca inspection

Cloaca inspections were performed on 47 female and 61 male breeding shearwaters sexed by molecular tests. The cloacal size distribution was bimodal (Fig. 3). Mean female cloacal width (5.6 mm, range 4–8 mm) was 1.9 mm wider (95% CI 1.6–2.3 mm, $t_{106} = 11.5$, $P < 0.001$) than mean male cloacal width (3.7 mm, range 2–6 mm). Linear discriminant function analysis indicated a sex classification cut-point at 4.7 mm cloacal width.

Visual assessments of cloacal size sexed 46 of 47 females (98%) correctly, which was significantly greater than 47 of 61 (77%) for males ($\chi^2_1 = 9.5$, $P = 0.002$; Table 1). This difference resulted from a decision to classify medium-sized cloacae as females (Fig. 3). Errors from cloacal width measurements and discriminant function analysis were more nearly balanced, with 42 of 47 females (89%) and 51 of 61 males (84%) sexed correctly. Overall accuracy, for males and females combined, was 93 of 108 (86%, 95% CI 76–96%) for both cloacal size and width data.

Female cloacal size was larger than that of the male in 36 of 39 breeding pairs (92%), or 35 of 39 (90%) by cloacal width measurements (Table 1). There was one indeterminate pair with "medium" cloacal size descriptions for both birds, and there were three pairs with equal width measurements. Identical female and

TABLE 3
Within-pair comparisons of Wedge-tailed Shearwater biometrics (male size > female size)

Biometric	Observed pairs proportion			Bootstrap random pairs proportion			Bootstrap (observed − random)	
	n pairs	M > F (%)	95% CI (%)	n pairs	M > F (%)	95% CI[a] (%)	Diff.[b] (%)	95% CI[c] (%)
Wing	39	62	46–75	39	62	46–75	–0.6	–23 to 21
Tail	39	33	21–49	38	40	26–55	–6	–27 to 15
Head	39	87	73–94	39	84	70–93	4	–13 to 21
Culmen	39	77	62–87	39	77	62–87	0.2	–18 to 21
Nalospi	39	72	56–83	39	76	62–87	–4	–23 to 15
Bill depth	39	82	67–91	39	82	67–91	0.6	–18 to 18
Min. bill depth	39	87	73–94	39	82	67–91	5	–10 to 21
Bill width	38	76	61–87	37	69	51–80	8	–12 to 29
Tarsus	39	69	54–81	39	65	48–77	5	–18 to 26
Mid-toe & claw	37	62	46–76	37	64	49–78	–2	–24 to 21
Mass	38	45	30–60	36	44	29–59	1.0	–23 to 24
Discriminant score DS_1	39	92	80–97	39	89	76–96	3	–8 to 15

[a] Bootstrap random female-male pairs proportions were estimated from the entire biometrics dataset (Table 2) and are reported with Wilson score confidence intervals.
[b] Biases in the bootstrap observed pairs proportions were –0.4 to +0.3% and carry over to bootstrap (observed pairs – random pairs) differences.
[c] Bootstrap differences reported with percentile confidence intervals.

male measurements within-pairs were classified as failures because indeterminate results are no more useful than a wrong result. Within-pair comparisons correctly reclassified five males having medium cloacal size descriptions. Within-pair comparison accuracy was 13% greater (95% CI 0.8–26%) than accuracy from visual assessments of cloacal size for the 39 males (79%) in the pairs sample (McNemar mid-*P* test X = 5, n = 5, *P* = 0.03; Table 1).

Biometrics

The biometrics sample included 47 female and 51 male shearwaters sexed by molecular tests. Males averaged significantly larger than females for all measurements ($t \geq 2.6$, $df \geq 94$, $P \leq 0.012$) except tail length and mass (Table 2). Maximum male-biased SSD was 5–6% for bill depth measurements. Mean SSD excluding tail length and mass was 3%.

A biometric discriminant function combining bill depth and total head length sexed 38 of 47 females (81%, 95% CI 67–90%) and 41 of 51 males (80%, 95% CI 68–89%) correctly. Overall accuracy was 79 of 98 birds (81%, 95% CI 72–87%). The discriminant function obtained was:

$DS_1 = 1.371$ (bill depth) + 0.401 (total head length) − 51.495

where a bird is classified as male when discriminant score $DS_1 >$ and female when $DS_1 < 0$ (Fig. 4).

Male total head length and minimum bill depth were larger than for the female in 34 of 39 breeding pairs (87%; Table 3). Male DS_1 was larger than the female in 36 of 39 pairs (92%). This second result was 10% greater (95% CI −3 to 25%) than DS_1 accuracy for the 39 females (82%) in the pairs sample and 8% greater (95% CI −5 to 22%) than DS_1 accuracy for the 39 males (85%) in the pairs sample. These moderate improvements in sex classification accuracy for within-pair comparisons over DS_1 applied to individuals were not significant (McNemar mid-*P* tests, $P \geq 0.13$). Statistical tests for paired proportions have low power when the sum of discordant paired results is small. Larger sample sizes will tend to increase the numbers of discordant results and statistical power.

Within-pair comparisons of biometrics from random female-male pairs agreed closely with those from observed breeding pairs (Table 3). All 95% confidence intervals for bootstrap (observed pairs − random pairs) differences in proportions included zero. There were no strong, positive within-pair correlations between female and male biometrics to suggest size-assortative mating ($r_S \leq 0.27$, n = 37–39 pairs, $P \geq 0.10$). The within-pair correlation for overall body size (first principal component computed for wing, head, bill depth and tarsus measurements) was $r_s = -0.11$ (n = 38 pairs, *P* = 0.52).

Acoustics

The acoustic measurements sample included 471 burrow call recordings from 45 female and 60 male shearwaters sexed by molecular tests. To remove pseudo-replication (multiple calls per individual), the discriminant function analysis used individual means (Totterman 2014).

An acoustic discriminant function combining syllable two, note two fundamental frequency and note length sexed 44 of 45 females (98%, 95% CI 88–100%) and 58 of 60 males (97%, 95% CI 89–99%) correctly. Overall accuracy was 102 of 105 birds (97%, 95% CI 92–99%). The discriminant function obtained was:

$DS_2 = 0.0618$ (note 2-2 frequency) + 2.901 (note 2-2 length) − 16.927

where a bird is classified as male when discriminant score $DS_2 > 0$ and female when $DS_2 < 0$ (Fig. 5a). Evaluated on the 471 calls, DS_2 sexing accuracy was maintained at 98% for females and 94% for males (Fig. 5b).

Male mean syllable two, note one frequency was higher than the female mean in 36 of 38 breeding pairs (95%; Table 4). Male DS_2 was larger than the female in 100% of pairs. The small improvements in sex classification accuracy for within-pair comparisons of DS_2 over DS_2 classification of individual females and males were not significant (McNemar mid-*P* tests, $P \geq 0.25$).

Within-pair comparisons of acoustic measurements from random female-male pairs agreed closely with those from observed breeding

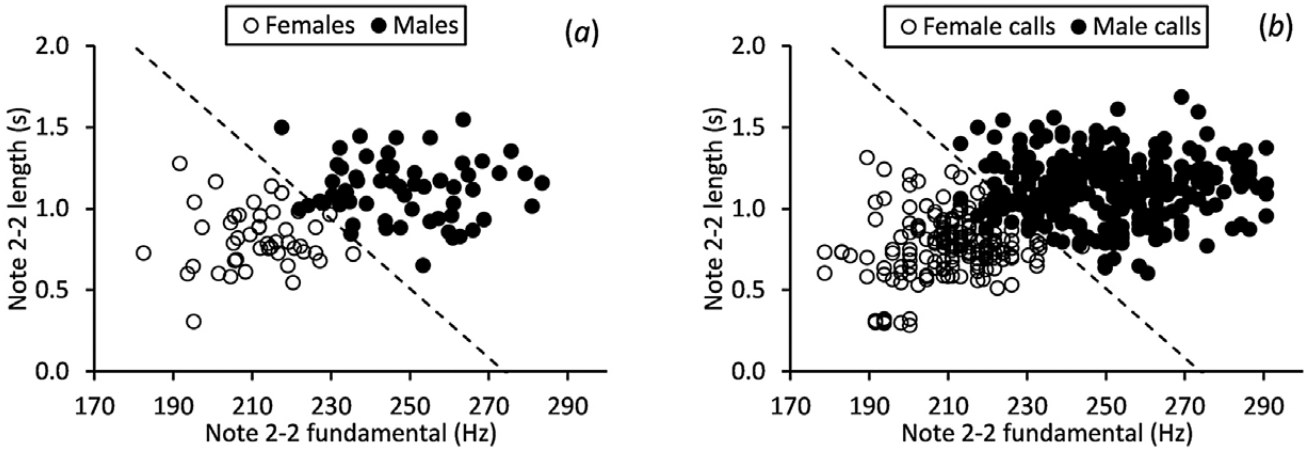

Fig. 5. Syllable two, note two fundamental frequency and note length for Wedge-tailed Shearwaters: (a) individual means (n = 45 females, 60 males); and (b) calls (n = 471). Dashed lines indicates equal female/male probability from the acoustic discriminant function ($DS_2 = 0$). Points plotting above the lines ($DS_2 > 0$) are classified as males and points below ($DS_2 < 0$) as females.

pairs (Table 4). All 95% confidence intervals for bootstrap (observed pairs − random pairs) differences in proportions included zero (Table 4). There were no strong, positive within-pair correlations to suggest assortative mating for those acoustic variables measured. Most within-pair correlations were negative ($r_S \leq 0.02$), including a modest correlation for syllable one, note one frequency ($r_S = -0.47$, $P = 0.003$).

Compared with other field sexing methods, the 97% overall DS_2 accuracy was 11% greater (95% CI 4–18%) than the 86% for cloacal sexing ($\chi^2_1 = 8.3$, $P = 0.004$) and 17% greater (95% CI 8–25%) than the 81% for biometric discriminant function DS_1 ($\chi^2_1 = 14.3$, $P < 0.001$).

Playback-response tests

Playback-response tests were performed for 21 female and 29 male shearwaters sexed by molecular tests. Eighteen of 21 females (86%, 95% CI 65–95%) responded exclusively to female call playbacks or were silent, which was 14% fewer (95% CI 0.7–29%) than 29 of 29 males (100%, 95% CI 88–100%) responding to male calls ($\chi^2_1 = 4.3$, $P = 0.04$). Fourteen of 21 females (67%, 95% CI 45–83%) called back, and seven were silent. This response rate was 33% lower (95% CI 13–53%) than for 29 of 29 males (100%) responding to male playbacks ($\chi^2_1 = 11.0$, $P < 0.001$) as well as 30% lower (95% CI 9–51%) than for 28 of 29 males (97%, 95% CI 83–100%) responding to female playbacks ($\chi^2_1 = 7.9$, $P = 0.005$). Overall playback-response sexing accuracy was 47 of 50 birds (94%, 95% CI 84–98%) and the overall response rate was 43 of 50 birds (86%, 95% CI 74–93%).

The 14% error rate in females (above) agreed with the results of my previous study, in which two of 18 females (11%) responded to stranger male playbacks in the vocal sex recognition experiment (Totterman 2014; $\chi^2_1 = 0.09$, $P = 0.77$). The 67% female response rate in this study was 27% lower (95% CI 5–49%) than the results of my previous study (Totterman 2014), in which 28 of 30 females (93%) responded to stranger female playbacks, when female and male calls were played on different nights ($\chi^2_1 = 5.9$, $P = 0.01$). There were no significant differences among males in responsiveness to playback of stranger female and male calls in these two experiments ($\chi^2_1 \leq 0.45$, $P \geq 0.50$). Eleven females and

TABLE 4
Within-pair comparisons of Wedge-tailed Shearwater burrow call acoustic measurements (n = 38 pairs)

Acoustic measurement	Within-pair	Observed pairs		Bootstrap random pairs		Bootstrap (observed − random)	
		Prop. (%)	95% CI (%)	Prop. (%)	95% CI[a] (%)	Diff.[b] (%)	95% CI[c] (%)
Note 1-1 fundamental	M > F	92	79–97	95	83–99	−3	−13 to 11
Note 1-2 fundamental	M > F	92	79–97	93	79–97	−1.2	−13 to 11
Note 2-1 fundamental	M > F	95	83–99	94	83–99	0.3	−11 to 11
Note 2-2 fundamental	M > F	92	79–97	99	87–100	−7	−16 to 3
Note 3-1 fundamental	M > F	89	76–96	91	79–97	−2	−16 to 11
Note 3-2 fundamental	M > F	92	79–97	95	83–99	−3	−13 to 8
Note 1-1 length	M > F	82	67–91	83	70–93	−2	−18 to 16
Note 1-2 length	M > F	84	70–93	88	73–94	−4	−18 to 11
Note 2-1 length	M > F	82	67–91	82	67–91	−0.3	−18 to 18
Note 2-2 length	M > F	84	70–93	87	73–94	−3	−18 to 13
Note 3-1 length	M > F	84	70–93	82	67–91	2	−16 to 18
Note 3-2 length	M > F	87	73–94	87	73–94	0.2	−16 to 16
Syllable 1 note interval	F > M	71	55–83	68	53–81	3	−18 to 24
Syllable 2 note interval	F > M	61	45–74	61	45–74	−3	−24 to 21
Syllable 3 note interval	F > M	71	55–83	63	47–77	8	−13 to 29
Syllable 1-2 interval	M > F	63	47–77	58	42–72	5	−16 to 26
Syllable 2-3 interval	M > F	63	47–77	60	45–74	2	−21 to 24
Number of syllables	F > M	55	40–70	56	40–70	−0.4	−24 to 21
Crescendo peak syllable	F > M	74	58–85	64	47–77	10	−11 to 32
Discriminant score DS_2	M > F	100	91–100	100	91–100	0.2	0.0 to 3

[a] Bootstrap random female-male pairs proportions were estimated from the entire acoustics dataset and are reported with Wilson score confidence intervals.

[b] Biases in the bootstrap observed pairs proportions were −2.7 to +0.4% and carry over to bootstrap (observed pairs − random pairs) differences.

[c] Bootstrap differences reported with percentile confidence intervals.

24 males in the second experiment were not tested in the first, and the independent proportions "N – 1" *chi*-squared test is appropriate for these comparisons.

Compared with other field sexing methods, overall sexing accuracy of playback-response was 8% greater (95% CI −1.4 to 17%) than cloaca inspection (χ^2_1 = 2.1, P = 0.15; not significant given the sample sizes in this study) and 13% greater (95% CI 3–24%) than the biometric discriminant function DS_1 (χ^2_1 = 4.7, P = 0.03).

DISCUSSION

Cloacal sexing

Cloaca inspection is a simple field method for sexing breeding petrels that provides immediate results. Its major limitation is that the female cloaca gradually relapses to its normal form after laying (Serventy 1956, Boersma & Davies 1987). Birds in the present study were examined from about one to four weeks after laying, and overall cloacal sexing accuracy was 86%. Roberts *et al.* (1974) also examined Wedge-tailed Shearwaters and commented that cloacal sexing was less reliable from two weeks after laying. Boersma & Davies (1987) examined breeding Fork-tailed Storm-Petrels *Oceanodroma furcata* from one to four weeks after laying and reported 93% accuracy (n = 150 females, 150 males). O'Dwyer *et al.* (2006) examined Gould's Petrels *Pterodroma leucoptera* during the first half of the incubation period and reported 96% accuracy (n = 54 females, 74 males). Within-pair comparisons of cloacal size for Wedge-tailed Shearwaters resolved some errors in individual assessments, and sexing accuracy was 92% after these comparisons, similar to the 96% (n = 150 pairs) reported for Fork-tailed Storm-Petrels (Boersma & Davies 1987).

Cloaca inspection requires careful handling because struggling birds can present extruded cloacae (O'Dwyer *et al.* 2006). Three male shearwaters in this study presented a "large" cloaca. No females presented a "small" or "extra-small" cloaca. Intermediate assessments of cloacal size (i.e. "medium" or "moderate") are ambiguous. Caliper measurements are more objective, but can lack other information, including prominence of the cloaca lips and colour. The sex classification cut-point will also vary, gradually decreasing after laying.

Subjective assessments can be affected by observer-expectation bias, i.e. when an observer's expectations or wishes influence results (Balph & Romesburg 1986). Visual assessments of cloacal size and shape can be influenced by observations preceding cloaca inspection (e.g. biometrics and behaviour), when a sample includes breeding pairs (i.e. presuming the complementary sex for the second bird assessed) and when the sex-ratio of birds sampled is skewed (i.e. when the observer is expecting a nearly balanced number of females and males). Applying a strict dichotomous classification avoids ambiguous cloacal size descriptions noted above, but encourages the observer to consider other information when forced to decide between "large" or "small" for those intermediate sizes.

Biometric sexing

Male Wedge-tailed Shearwaters from Muttonbird Island averaged 3% larger in size than females, which was similar to 2% for museum specimens from 25 breeding localities (Bull *et al.* 2005) and 2% from Hawaii (Shallenberger 1973). Bill depth was the most sexually dimorphic: 5% and 6% in this study, 4% in Bull *et al.* (2005) and 4% in Shallenberger (1973).

The overall 81% biometric discriminant function sexing accuracy for Wedge-tailed Shearwaters in this study was similar to the 83% (n = 197 females, 193 males) reported for Short-tailed Shearwaters *P. tenuirostris* by Carey (2011). Einoder *et al*. (2008) reported 92% accuracy for Short-tailed Shearwaters, although sample sizes were small (n = 26 females, 25 males) and that estimate has low precision. Other studies have reported 90% accuracy (n = 52 females, 50 males) for Flesh-footed Shearwater *P. carneipes* (Thalmann *et al*. 2007), 90% (n = 20 females, 32 males) for Balearic Shearwater *P. mauretanicus* (Genovart *et al*. 2003) and 87% (n = 42 females, 44 males) for Yelkouan Shearwater *P. yelkouan* (Bourgeois *et al*. 2007). Bill depth and total head length were selected in five of these six studies (only Bourgeois *et al*. 2007 did not measure total head length). These two measurements should be sufficient for biometric sexing of large shearwaters.

Two problems for biometric discriminant function sexing are geographical variation in size between colonies and between-observer variation in measurements. Einoder *et al.* (2008) reported that sexing accuracy decreased from 92% to 70% when applying a discriminant function from one Short-tailed Shearwater colony to another colony. For Flesh-footed Shearwaters, accuracy decreased from 90% to 69% when a single-colony discriminant function was tested on a fishing bycatch sample (Thalmann *et al.* 2007). Van Franeker & Ter Braak (1993) described *ad hoc* methods for adjusting discriminant functions when there is geographical variation in size.

Bill depth was the most effective single biometric for predicting sex in Wedge-tailed Shearwaters; however, the 0.6 mm (male−female) mean difference is small. Between-observer variation for bird measurements can exceed within-observer variation (e.g. Barrett *et al.* 1989, Goodenough *et al.* 2010), and any measurement biases can reduce discriminant function accuracy (Francis & Mattlin 1986). Therefore, before applying biometric discriminant functions, other observers should check that their measurements are consistent with means and variances of the data used to create the discriminant function (Lorentsen & Røv 1998).

Within-pair comparisons of minimum bill depth improved biometric sexing accuracy for Wedge-tailed Shearwaters to 87%. This result is comparable to the 84% accuracy (n = 25 pairs) reported for within-pair bill depth comparisons in Yelkouan Shearwaters (Bourgeois *et al*. 2007) and the 92% accuracy (n = 171 pairs) reported for Short-tailed Shearwaters (Carey 2011). Genovart *et al*. (2003) reported 100% success for within-pair bill depth comparisons in Balearic Shearwaters, although their sample size was only 10 pairs. Within-pair comparisons should not be affected by geographical variation in size and, for a single measurer, between-observer variation. However, within-pair comparisons are limited to breeding pairs in which both birds have been assessed.

Whether size-assortative mating occurs in shearwaters from the genus *Puffinus* is unclear. Size-assortative mating was not detected for Wedge-tailed Shearwaters from Muttonbird Island. Einoder *et al.* (2008) reported assortative mating with respect to bill depth for Short-tailed Shearwaters from one colony, but Carey (2011) found no evidence for size-assortative mating from another colony

in a study involving much larger sample sizes. SSD is weak in shearwaters from the genus *Puffinus* (Bull *et al.* 2005), and visual communication at breeding colonies at night is limited to close contact and large-scale gestures (Shallenberger 1973). A mechanism for evaluation of subtle differences in size is unclear. A simple explanation for why bill depth has been relatively accurate for within-pair classification of sexes is that bill depth is the most sexually dimorphic biometric in shearwaters.

Acoustic sexing

Although Wedge-tailed Shearwaters do not have distinct female and male calls (Totterman 2014), the acoustic discriminant function combining syllable two, note two frequency and length sexed 97% of birds correctly. Brooke (1988) similarly proposed a combination of frequency and syllable length to classify female and male Great Shearwaters *P. gravis*. He reported 100% accuracy, although his sample sizes were small (n = 10 females, 12 males). Four other shearwaters from the genus *Puffinus* studied have distinct female and male voices. Although not evaluated for known-sex birds, acoustic sexing by ear was apparently 100% accurate for Manx Shearwater *P. puffinus* (Brooke 1978), Audubon's Shearwater *P. lherminieri* (Bretagnolle *et al.* 2000) and Little Shearwater *P. assimilis* (James & Robertson 1985). Bourgeois *et al*. (2007) were able to compare acoustic sexing with molecular sexing results and reported 100% accuracy (n = 6 females, 10 males) for Yelkouan Shearwaters.

Exhalant notes are emphasised in Wedge-tailed Shearwater burrow calls, whereas discriminant function analysis selected an inhalant note for acoustic sexing. Shallenberger (1973) suggested that inhalant notes were more reliable because birds exercise less vocal control during inhalation. These results do not imply that Wedge-tailed Shearwaters use inhalant notes for sex recognition, however. Playback experiments using modified or synthetic calls are required to identify which call properties the birds are sensitive to (e.g. Taoka & Okumura 1990).

Two challenges for acoustic sexing are stimulating birds to call and taking into account geographical variation. Playback response rates at night are higher than during the day (Shallenberger 1973, Burger & Lawrence 2001). Female burrow call playbacks were effective for eliciting call responses from both female and male Wedge-tailed Shearwaters (Totterman 2014). Geographical variation in calls is expected (Bretagnolle 1996) and requires development of colony-specific acoustic discriminant functions.

Playback-response sexing

Selectivity towards same-sex call playbacks has been demonstrated for a variety of petrels (e.g. Brooke 1978, Brooke 1988, Taoka *et al.* 1989a, Taoka *et al.* 1989b, Cure *et al.* 2009, Totterman 2012). However, no playback-response sexing methods for petrels could be found in my literature searches.

Overall playback-response sexing accuracy for Wedge-tailed Shearwaters was 94%, which is slightly less than the 97% accuracy yielded by the acoustic discriminant function. Playback-response tests are simpler than field recording and acoustic analysis, and results are available immediately. Local call recordings can be used to take into account geographical variation. Similar playback-response methods could easily be developed for other petrels in which females and males are selective towards same-sex calls.

Two challenges for playback-response tests in Wedge-tailed Shearwaters are the 14% error rate and the 33% non-response rate in females. Owing to overlapping female and male call parameters (Totterman 2014) and environmental effects on sound propagation (Jouventin & Aubin 2000), females could have misidentified the sex of stranger male playbacks on a few occasions. Two of three females that responded to a male call in the playback-response sexing experiment did not respond to different male calls in repeat tests performed on subsequent nights.

Females may be reluctant to respond to stranger female calls following male playbacks. Increasing the number or duration of female playbacks might not improve response rates (pers. obs.). Reversing the playback sequence (female calls first) would be more effective for exciting females but could result in their responding to subsequent male calls (e.g. Taoka *et al.* 1989a). Playing male and female calls on different nights is inefficient. Other researchers have waited for five or 15 minutes between playing back calls with different sexual or individual identities (Jouventin & Aubin 2000, Cure *et al.* 2009). Similar intervals could also make Wedge-tailed Shearwater responses to male and then female call playbacks more independent, thereby increasing female response rates.

CONCLUSIONS

Biometric sexing performs poorly when levels of SSD are low. Researchers should weigh the value of biometric data against the time spent measuring birds. They might select a minimal set of biometrics for specific research objectives. Cloaca inspection is a good "quick look" method for sexing breeding petrels, but with variable accuracy. Even a highly accurate result such as 96% (O'Dwyer *et al.* 2006) is equivalent to one cloacal sexing error in 25, which can result in statistical outliers. Field sexing methods that have imperfect accuracy should be restricted to field use.

For definitive results, feathers are easy to collect and *CHD*-based molecular sexing techniques are affordable and objective (Dubiec & Zagalska-Neubauer 2006). Molecular tests can be applied to birds of all ages, breeding status and in all seasons. Body moult in adult Wedge-tailed Shearwaters occurs in summer (Swanson & Merritt 1974), and there should be no ethical concerns about plucking a few breast feathers from birds during the breeding season. Blood is recommended for more detailed molecular studies requiring larger quantities of high-quality DNA and for archival purposes (McDonald & Griffith 2011).

Wedge-tailed Shearwaters primarily use vocalisations to communicate in the dark (Shallenberger 1973), and individual birds are sensitive to sexual differences in burrow calls (Totterman 2014). Acoustic and playback-response field sexing methods for Wedge-tailed Shearwaters were more accurate than cloaca inspection and biometric methods in this study. For most burrows, acoustic and playback-response methods could simultaneously determine occupancy and sexual identity of the occupant without physical intervention. Further investigations of the vocal behaviour of petrels and development of practical field applications are encouraged.

ACKNOWLEDGEMENTS

Work on Muttonbird Island was permitted under Scientific License 101052, authorised by the New South Wales (NSW) Office of Environment and Heritage, and Animal Research Authority

12/4839, authorised by the NSW Department of Primary Industries. I am thankful to Ann Walton, Coffs Coast Ranger, and Glenn Storrie, Area Manager, National Parks and Wildlife Service, for their support and helpful advice. Deanne Waine from The University of Queensland and Dr. Richard McCoy from Gribbles Veterinary are thanked for their efficient genetics laboratory work and support. Colin Miskelly and Bo Totterman provided helpful pre-submission reviews of this paper.

REFERENCES

BALPH, D.F. & ROMESBURG, H.C. 1986. The possible impact of observer bias on some avian research. *Auk* 103: 831–832.

BARTLE, J.A. 1968. Observations on the breeding habits of Pycroft's petrel. *Notornis* 15: 70–99.

BARRETT, R.T., PETERZ, M., FURNESS, R.W. & DURNICK, J. 1989. The variability of biometric measurements. *Ringing and Migration* 10: 13–16.

BOERSMA, P.D. & DAVIES, E.M. 1987. Sexing monomorphic birds by vent measurements. *Auk* 104: 779–783.

BOURGEOIS, K., CURE, C., LEGRAND, J., GOMEZ-DIAZ, E., VIDAL, E., AUBIN, T. & MATHEVON, N. 2007. Morphological versus acoustic analysis: what is the most efficient method for sexing yelkouan shearwaters *Puffinus yelkouan*? *Journal of Ornithology* 148: 261–269.

BRETAGNOLLE, V. 1996. Acoustic communication in a group of nonpasserine birds, the petrels. In: Kroodsma, D.E. & Miller, E.H. (Eds). *Ecology and evolution of acoustic communication in birds*. Ithaca, NY: Cornell University Press. pp. 160–178.

BRETAGNOLLE, V., ATTIÉ, C. & MOUGEOT, F. 2000. Audubon's Shearwaters *Puffinus lherminieri* on Réunion Island, Indian Ocean: behaviour, census, distribution, biometrics and breeding biology. *Ibis* 142: 399–412.

BROOKE, M. de L. 1978. Sexual differences in the voice and individual vocal recognition in the Manx shearwater (*Puffinus puffinus*). *Animal Behaviour* 26: 622–629.

BROOKE, M. de L. 1988. Sexual dimorphism in the voice of the Greater Shearwater. *Wilson Bulletin* 100: 319–323.

BULL, L.S., BELL, B.D. & PLEDGER, S. 2005. Patterns of size variation in the shearwater genus *Puffinus*. *Marine Ornithology* 33: 27–39.

BURGER, A.E. & LAWRENCE, A.D. 2001. Census of Wedge-tailed Shearwaters *Puffinus pacificus* and Audubon's Shearwaters *P. lherminieri* on Cousin Island, Seychelles using call-playback. *Marine Ornithology* 29: 57–64.

CAMPBELL, I. 2007. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine* 26: 3661–3675.

CAREY, M.J. 2011. Sexual size dimorphism, within-pair comparisons and assortative mating in the short-tailed shearwater (*Puffinus tenuirostris*). *Notornis* 58: 8–16.

COPESTAKE, P.G., CROXALL, J.P. & PRINCE, P.A. 1988. Use of cloacal sexing techniques in mark-recapture estimates of breeding population size in Wilson's Stormpetrel *Oceanites oceanicus* at South Georgia. *Polar Biology* 8: 271–279.

CURE, C., AUBIN, T. & MATHEVON, N. 2009. Acoustic convergence and divergence in two sympatric burrowing nocturnal seabirds. *Biological Journal of the Linnean Society* 96: 115–134.

DECHAUME-MONCHARMONT, F., MONCEAU, K. & CEZILLY, F. 2011. Sexing birds using discriminant function analysis: a critical appraisal. *Auk* 128: 78–86.

DUBIEC, A. & ZAGALSKA-NEUBAUER, M. 2006. Molecular techniques for sex identification in birds. *Biological Letters* 43: 3–12.

EINODER, L.D., PAGE, B. & GOLDSWORTHY, S.D. 2008. Sexual size dimorphism and assortative mating in the Short-tailed Shearwater *Puffinus tenuirostris*. *Marine Ornithology* 36: 167–173.

FAGERLAND, M.W., LYDERSEN, S. & LAAKE, P. 2013. The McNemar test for binary matched-pairs data: mid-*p* and asymptotic are better than exact conditional. *BMC Medical Research Methodology* 13: 91.

FRANCIS, R.I.C.C. & MATTLIN, R.H. 1986. A possible pitfall in the morphometric application of discriminant analysis: measurement bias. *Marine Biology* 93: 311–313.

GENOVART, M., MCMINN, M. & BOWLER, D. 2003. A discriminant function for predicting sex in the Balearic Shearwater. *Waterbirds* 26: 72–76.

GOODENOUGH, A.E., STAFFORD, R., CATLIN-GROVES, C.L., SMITH, A.L. & HART, A.G. 2010. Within- and among-observer variation in measurements of animal biometrics and their influence on accurate quantification of common biometric-based condition indices. *Annale Zoologici Fennici* 47: 323–334.

GRIFFITHS, R., DOUBLE, M.C., ORR, K. & DAWSON, R.J. 1998. A DNA test to sex most birds. *Molecular Ecology* 7: 1071–1075.

JAMES, P.C. & ROBERTSON, H.A. 1985. Sexual dimorphism in the voice of the Little Shearwater *Puffinus assimilis*. *Ibis* 127: 388–390.

JOUVENTIN, P. & AUBIN, T. 2000. Acoustic convergence between two nocturnal burrowing seabirds, experiments with a penguin *Eudyptula minor* and a shearwater *Puffinus tenuirostris*. *Ibis* 142: 645–656.

LACHENBRUCH, P.A. & MICKEY, M.R. 1968. Estimation of error rates in discriminant analysis. *Technometrics* 10: 1–11.

LORENTSEN, S & RØV, N. 1998. Sex determination of Antarctic Petrels *Thalassoica antarctica* by discriminant analysis of morphometric characters. *Polar Biology* 14: 143–145.

MACKIN, W.A. 2005. Neighbor-stranger discrimination in Audubon's shearwater (*Puffinus l. lherminieri*) explained by a "real enemy" effect. *Behavioural Ecology and Sociobiology* 59: 326–332.

MCDONALD, P.G. & GRIFFITH, S.C. 2011. To pluck or not to pluck: the hidden ethical and scientific costs of relying on feathers as a primary source of DNA. *Journal of Avian Biology* 42: 197–203.

NEWCOMBE, R.G. 1998a. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine* 17: 857–872.

NEWCOMBE, R.G. 1998b. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 17: 873–890.

NEWCOMBE, R.G. 1998c. Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* 17: 2635–2650.

NORRIS-CANEDA, K.H. & ELLIOTT, J.D. 1998. Sex identification in raptors using PCR. *Journal of Raptor Research* 32: 278–280.

O'DWYER, T.W, PRIDDEL, D., CARLILE, N., BARTLE, J.A. & BUTTEMER, W.A. 2006. An evaluation of three field techniques for sexing Gould's Petrels (*Pterodroma leucoptera*) (Procellariidae). *Emu* 106: 245–252.

PEARSON, E.S. 1947. The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table. *Biometrika* 34: 139–167.

R DEVELOPMENT CORE TEAM. 2013. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. [Available online at: http://www.R-project.org/; accessed 30 July 2013].

RISING, J.D. & SOMERS, K.M. 1989. The measurement of overall body size in birds. *Auk* 106: 666–674.

ROBERTS, P.E., MERRITT, F.D. & FLOYD, R.B. 1974. Wedge-tailed Shearwaters on Muttonbird Island, Coffs Harbour, NSW. *Emu* 75: 19–22.

ROBERTSON, B.C. & GEMMELL, N.J. 2006. PCR-based sexing in conservation biology: wrong answers from an accurate methodology? *Conservation Genetics* 7: 267–271.

SERVENTY, D.L. 1956. A method of sexing petrels in field observations. *Emu* 56: 213–214.

SHALLENBERGER, R.J. 1973. Breeding biology, homing behaviour, and communication patterns of the Wedge-tailed Shearwater, *Puffinus pacificus chlororhynchus* [PhD dissertation]. Los Angeles, CA: University of California, Los Angeles.

SWANSON, N.M. & MERRITT, F.D. 1974. The breeding cycle of the Wedge-tailed Shearwater on Mutton Bird Island, N.S.W. *Australian Bird Bander* 12: 3–9.

TAOKA, M., WON, P. & OKUMURA, H. 1989a. Vocal behaviour of Swinhoe's Storm-Petrel (*Oceanodroma monorhis*). *Auk* 106: 471–474.

TAOKA, M., SATO, T., KAMADA, T. & OKUMURA, H. 1989b. Sexual dimorphism of chatter-calls and vocal sex recognition in Leach's Storm-Petrels (*Oceanodroma leucorhoa*). *Auk* 106: 498–500.

TAOKA, M. & OKUMURA, H. 1990. Sexual differences in flight calls and the cue for vocal sex recognition of Swinhoe's Storm-Petrels. *Condor* 92: 571–575.

THALMANN, S., BAKER G.B., HINDELL, M., DOUBLE, M.C. & GALES, R. 2007. Using biometric measurements to determine gender of Flesh-footed Shearwaters, and their application as a tool in long-line by-catch management and ecological field studies. *Emu* 107: 231–238.

THOMPSON, N.S., LEDOUX, K. & MOODY, K. 1994. A system for describing bird song units. *Bioacoustics* 5: 267–279.

TOTTERMAN, S.L. 2012. Sexual differences in vocalisations and playback-response behaviour of the Vanuatu petrel (*Pterodroma occulta*). *Notornis* 59: 97–104.

TOTTERMAN, S.L. 2014. Sexual and individual differences in wedge-tailed shearwater (*Puffinus pacificus*) burrow calls and vocal recognition. *Notornis* 61: 121–130.

VAN FRANEKER, J.A. & TER BRAAK J.F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. *Auk* 110: 492–502.

VENABLES, W.N. & RIPLEY, B.D. 2002. *Modern Applied Statistics with S*. 4th ed. New York: Springer.

WARHAM, J. 1990. *The petrels: their ecology and breeding systems*. London: Academic Press.

WENZEL, D. & ZAPF, A. 2013. diffdepprop: calculates confidence intervals for two dependent proportions. [Available online at: http://CRAN.R-project.org/package=diffdepprop; accessed 30 June 2014]

WILSON, E.B. 1927. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* 22: 209–212.

ZUUR, A.F., IENO, E.N. & ELPHICK, C.S. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1: 3–14.