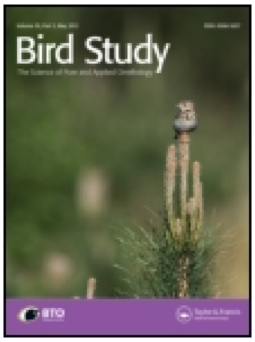

Bird Study

ISSN: 0006-3657 (Print) 1944-6705 (Online) Journal homepage: https://www.tandfonline.com/loi/tbis20

The field determination of body size and condition in passerines: a report to the British Ringing Committee

A.G. Gosler, J.J.D. Greenwood, J.K. Baker & N.C. Davidson

To cite this article: A.G. Gosler, J.J.D. Greenwood, J.K. Baker & N.C. Davidson (1998) The field determination of body size and condition in passerines: a report to the British Ringing Committee, Bird Study, 45:1, 92-103, DOI: 10.1080/00063659809461082

To link to this article: https://doi.org/10.1080/00063659809461082

Bird Study (1998) **45,** 92–103

The field determination of body size and condition in passerines: a report to the British Ringing Committee

ANDREW G. GOSLER[1]*, JEREMY J.D. GREENWOOD[2], JEFF K. BAKER[2] *and* NICHOLAS C. DAVIDSON[3], [1]*Edward Grey Institute of Field Ornithology, Department of Zoology, University of Oxford,South Parks Rd, Oxford OX1 3PS, UK,* [2]*British Trust for Ornithology, Nunnery Place, Thetford, Norfolk, IP24 2PU, UK and* [3]*Joint Nature Conservation Committee, Northminster House, Peterborough, Cambridgeshire, UK*

This paper assesses the best measure of body size for a range of passerine species, from data provided by British bird observatories. For each species we determine the proportion of variance in body mass explained by wing length, tarsus length, tail length and head length (head + bill) by multiple regression, while also accounting for variance due to two condition components of mass, measured as fat and muscle scores. Across 26 species ranging in mean mass from 5.3 g (Goldcrest) to 100.3 g (Blackbird), fat score was consistently either the best predictor of mass (14 species) or in the set of significant predictors (21 species); muscle score was the best predictor of mass in only one species and a significant predictor in nine. Amongst linear measures, wing length was consistently the best size predictor. The first principal component of size (PC1) based on the four linear measurements frequently explained more variance in mass than wing length alone but the improvement was generally small. Reproducibility of measurements was generally better for wing length than for other linear measures, and in some cases very much better. On the basis of these findings, we recommend the following: wing length is used to give the best general measure of body size within species of passerines; fat and muscle scores, together with total body mass, are recorded to assess individual condition.

Birds differ in mass chiefly because they differ in size, or condition, or both. Differences in size between species are of great significance in studies of evolution,[1] ecology[2,3] and physiology;[2,4] for many such studies (e.g. comparing across orders) mass gives a reasonable approximation to size because the variation in mass due to differences in size is small within species compared with the range of variation between species. However, ornithologists often require a measure of body size for studies of a single species or closely related group of species (which may differ little in size), so that mass alone may give a poor measure of size because of differences in condition. This is most graphically demonstrated by considering changes in mass of individual birds: annual variation in mass of more than 50% is common in migrants[5,6] and mass changes of 10% are typical in many passerines within a single day.[7] These changes clearly are not being caused by changes in body size; hence, some measure of size is required which is independent of mass.

By body size, we mean the size of the skeletal frame upon which soft tissues are supported. This might best be measured as skeletal mass, or its close correlate, lean dry mass.[8] In practice, these are impossible to measure in live birds so that some correlated but accessible alternative is needed. Several measures have been used in the past such as the lengths of wing, tarsus, or head, depending generally on the type of birds concerned. Some workers have argued that no single dimension

*Correspondence author.
Email: andrew.gosler@zoology.ox.ac.uk

can adequately reflect size so that measurements should be combined as the first principal component (PC1) in a principal components analysis (PCA).[9] This represents just that part of the variation in the various measurements that is correlated and thus reflects general body size. Nevertheless, for passerines, the measurement from carpal to tip of the closed wing (wing length) has long been recommended as a general measure of body size.[10,11] This is partly because it is easily measured and partly because it was found to be very strongly correlated with lean dry mass in a study of the Savannah Sparrow *Passerculus sandwichensis*.[8]

Before wing length is accepted as the principal measure of intraspecific body size, two questions should be addressed. First, we need to know that it is a good predictor of size in a range of species other than the Savannah Sparrow. Secondly, concerns have often been expressed over the reproducibility of wing measurements between observers.[12,13] Although this is of lesser concern for studies involving just one or a few observers working together, it may be more serious should data be collected by many observers, such as by a national ringing scheme. Recent studies indicate that if bird ringers receive intensive training in measurement, as in Britain, this is probably not a major problem.[14] Nevertheless, it is valuable to know whether other dimensions (and PC1), perhaps with greater reproducibility, might not be equally good predictors of overall body size.

We assume that a correlation between body mass and some linear dimension, such as wing length, arises because the latter reflects the size component of mass. In this paper we examine, for each of a range of passerine species, which of a series of measurements gives the highest partial correlation with body mass. Mass was corrected for condition by including fat and muscle measures within the analyses. We believe that this correction gives the best approximation to the lean dry mass that we can make for live birds. We then compare the quality of wing length with PC1 as measures of size. Finally we compare the reproducibility[14,15] of these various size and condition measures.

METHODS

Between 1989 and 1991, British bird observatories were invited by the Biometrics Working Group of the BTO Ringing Committee, to submit sets of measurements of passerines processed during routine operations. To standardize methods, observers were issued with standard recording sheets and detailed instructions on the measurements to be taken. Where the numbers of birds being caught was so large that taking all measurements would have resulted in unacceptable delays in processing birds, observers took only the measures of greatest priority.

In order of priority, the measurements requested were: (1) mass taken on a Pesola balance to 0.1 g if below 50 g and to 1 g if greater than this; (2) wing length (maximum chord) taken to the nearest 1 mm on a stopped rule;[16] (3) fat score in the tracheal pit recorded on the six-point scale described by Gosler;[17] (4) pectoral muscle score taken on the three-point scale described by Gosler;[18] (5) maximum tarsus length taken to the nearest 0.5 mm with the foot abutted on a stopped rule at right angles to the tarsus, and measured to the proximal point of the intertarsal joint (thus including part of the tibiotarsus); (6) minimum tarsus taken to 0.1 mm with a vernier calliper, as maximum tarsus but to the proximal point of the tarsometatarsus as described by Svensson[19] (Fig. 18b); (7) head length taken with a vernier calliper from the bill tip to the posterior centre of the cranium to 0.1 mm; (8) in 1990 and 1991, minimum tarsus was replaced by tail length taken to the nearest 1 mm with an unstopped rule.[19]

All analyses were carried out within species. To avoid including breeding birds in the analyses, we have concentrated on individuals believed to have been on migration. These included 3221 birds from the beginning of March to the end of June (1445 in May) and 3916 from the beginning of July to the end of December (1679 in October). As described above, analyses were undertaken chiefly by multiple regression. As examples of the data for a typical migratory species in which both size and fat are significant predictors, Fig. 1 shows a bivariate plot of mass against wing length by fat score classes for the Northern Wheatear and Fig. 2 shows these data identified by sex after removing the variance due to the differences in fat score that were shown in Fig. 1. Scientific names of species

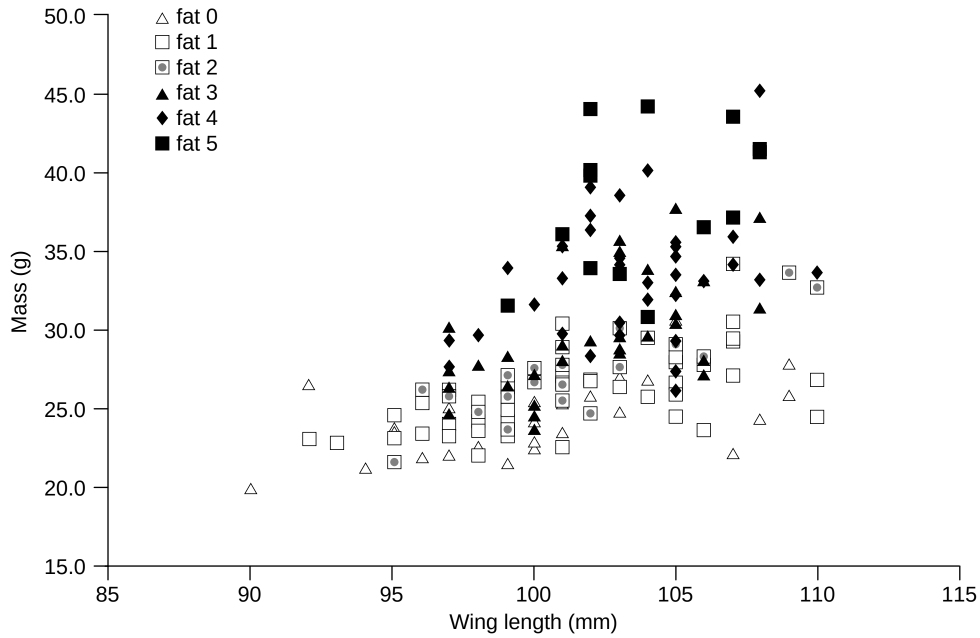

Figure 1. Body mass (g) plotted against wing length (mm), distinguished by fat score class, for 182 sexed Northern Wheatears trapped on migration at British Bird Observatories.

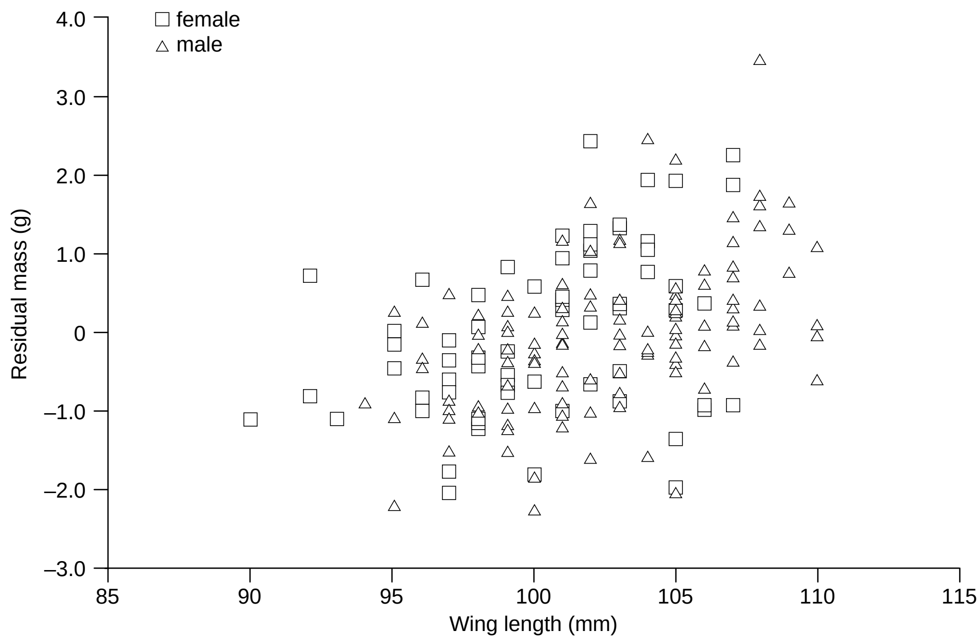

Figure 2. Residual body mass (g) after controlling for fat score class, plotted against wing length (mm), distinguished by sex, for 182 sexed Northern Wheatears trapped on migration at British Bird Observatories.

mentioned are given in Appendix 2.

Because many of the species concerned could not be sexed in the field, data from both sexes have been combined to allow comparison across as large a range of species as possible. Ideally sexes would have been separated because the relationship between mass and size may differ between sexes. However, in each species, the same samples were used to determine the best predictors of mass, so that this should not invalidate our results. This is also supported by plots such as Fig. 2 showing, for a highly dimorphic species (Northern Wheatear), a large overlap between sexes.

Because data from all measurements were not available for all birds, we have divided the whole data set into three inclusive groups. In total, data were available from 7137 birds of 90 species. However, only species with at least 12 observations were included in these analyses, reducing the total to 7008 of 51 species. All these birds (Group 1) had at least mass and wing length measured. Group 2 was a subset of Group 1 for which data for fat and muscle were also available (4899 birds of 39 species). Group 3 was a subset of Group 2 for which tarsus, tail and head measurements were also available (2317 birds of 26 species). In addition, a further 2266 measurements of Great Tits, collected by AGG at Wytham Woods, Oxon[17] were analysed as part of Groups 1 and 2. Species are presented in systematic order in the tables; a key to abbreviations used is given in Appendix 2.

Reproducibility shows the variation in measurements of the same bird taken by different observers. This differs from repeatability, which shows the variation in measurements taken by the same observer. To assess the reproducibility of different measurements, recording sheets were divided into two sections so that measurements taken by a visiting ringer and by an appointed 'standard' observer (usually the observatory warden) could be recorded for comparison. Visitors always measured first and the standard observer measured independently. Although some other protocol might have ensured greater independence of observations, this would have required birds to have been kept for longer than was acceptable for their welfare and would also have reduced greatly the available sample sizes. All standard observers were experienced instructors and, even if they knew the visitors' measurements, they were unlikely to be influenced by such knowledge. This view is strongly supported by the reproducibility data, which were rather low for some measurements. Following Gosler *et al.*[14] we calculated three components of reproducibility. Bias, the mean difference between measurements of two observers, measures the degree to which one observer is consistently longer or shorter than the other. This may be useful in cases where one type of observer such as trainees (visitors in the present case) tends to measure short.[14] Disagreement is the mean absolute value of these differences, i.e. ignoring the direction of the difference. Imprecision, calculated as the residual error mean square from a two-way ANOVA in which the factors were the individual bird and the class of observer (standard or visitor), represents the variation of repeat measurements after the species-specific bias (of visitors' measurements relative to standards) has been removed. Note that whereas bias and disagreement are simple linear measures given in millimetres (except in the case of fat and muscle scores), imprecision is a variance and has units of mm^2.

RESULTS AND DISCUSSION

Wing length as a predictor of mass

Figure 3 shows the distribution of correlations (r) between wing length and mass, which were significant in 32 (63%) of the 51 species in Group 1. The correlations were negative in four species, but none of these was significant (two-tailed tests). Across all 51 species, the mean correlation between wing length and mass was 0.34 ± 0.213 (sd, $n = 51$). In some cases, the lack of a correlation was due to a covariance between wing length and condition. For example, in the Dunnock, wing length and mass were not correlated ($r_{33} = 0.19$, ns) unless fat score was included in the analysis, when a significant proportion of the variance in mass was explained by wing length (Table 2).

To consider whether the relationship was improved generally by accounting for variation in body condition, as it was in the Dunnock, wing length, fat and muscle scores were entered as predictors in a stepwise multiple regression analysis for each species in Groups 2

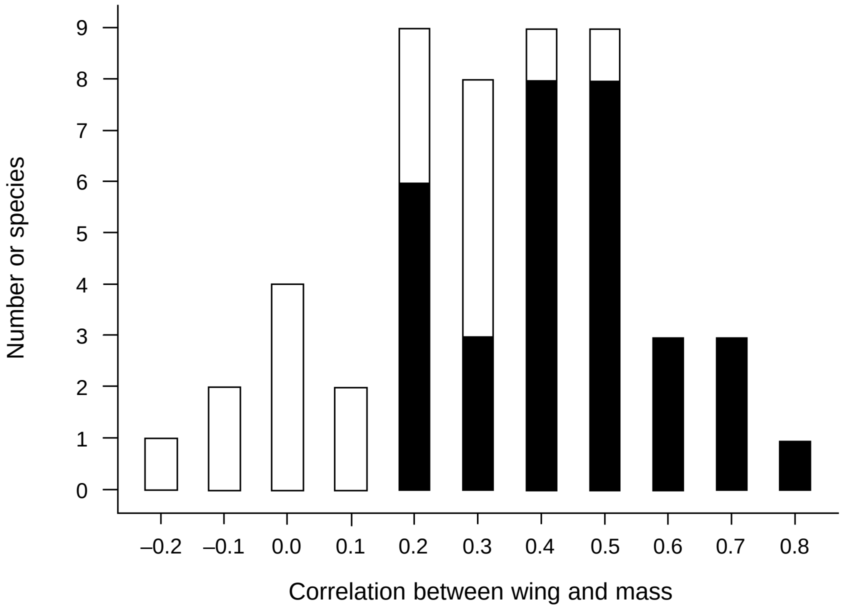

Figure 3. The distribution of correlations (r) between wing length and mass (■, $P < 0.05$; ❑, ns) across the 51 species in Group 1. Twelve species had sample sizes of fewer than 30 and, amongst these, only two were significant. Of 16 species with samples between 30 and 99, six correlations were not significant and in the remaining 23 species with samples greater than 100, only three were not statistically significant. The mean correlation (0.36 ± 0.187) for 39 species with 30 or more birds measured was little different from that of the whole data set (0.34 ± 0.217).

and 3. The full results of these analyses are presented in Tables 1 (for Group 2 species not in Group 3) and 2 (for Group 3 species). Across all 39 species the mean r^2 between wing and mass was 14.1 ± 13.73% (range 0–54%). When fat and muscle scores were also entered into the analyses, wing length explained on average 14.6 ± 13.51% (range 0–54%), a mean difference of only 0.4 ± 6.20%. Hence, although controlling fat and muscle made a significant difference in a few species such as the Dunnock, this was not generally so. There was no consistent effect, on the correlation between wing length and mass, of taking condition into account; changes in r^2 ranged from –19.0% (improved without fat and muscle) to +13.2% (improved with fat and muscle) but there was no change when it was averaged across species.

In these analyses, fat was selected as a significant predictor in 28 (72%) out of 39 species, with a mean partial r^2 of 26.4 ± 18.93% (n = 39). Muscle was selected as a significant predictor in 12 (31%) of 39 species with a mean r^2 of 7.8 ± 12.08 (n = 39). Thus variation in fat was generally the greatest cause of variation in mass. These three predictors together explained on average 41.6 ± 21.49% of the variance in mass (calculated for significant predictors only).

Correlations between fat and muscle scores were found in several species (Tables 1 and 2). These correlations were generally low (mean r = 0.25 ± 0.217), although 21 (54%) were significant (always positive). This is not surprising given that most of the birds were caught on migration, perhaps laying down fat not only in the tracheal pit but also within the muscles, as well as developing extra muscle tissue.[5] Furthermore, migrants that have exhausted their fat reserves may catabolize muscle tissue,[20] so birds with low muscle scores often also have reduced fat levels. The weak but significant correlation in the Wytham Great Tit data (Table 1) arises from parallel seasonal changes in fat and muscle. In Great Tits assessed over short periods, fat (which varies

Table 1. Results of stepwise multiple regression analyses of mass on wing length, fat and muscle scores. For each species (for abbreviations see Appendix 2) the table gives the percentage variance (independent r^2 values) in mass explained by each predictor and the total variance explained. Only values for significant predictors are shown. The final column indicates the correlation (r) between fat and muscle scores. The significance of the coefficient is indicated as $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. This table should be considered together with Table 2.

Species	n	*Wing*	*Fat*	*Muscle*	*Total*	r *F&M*
Sanma	13		35.0		35.0	–0.17 ns
Swall	33			19.4	19.4	0.35*
Wren	21	35.5			35.5	0.35 ns
Blare	16		21.8	54.0	75.8	0.37 ns
Whinc	17		47.3	20.9	68.2	0.15 ns
Rinou	12	9.4	72.3		81.6	0.23 ns
Field	29	12.9	35.8		48.6	0.11 ns
Barwa	14		41.0		41.0	0.43 ns
Greti	1157	25.4	16.4	1.1	42.9	0.13***
Rebsh	12		58.8		58.8	0.34 ns
Starl	21	54.0			54.0	0.01 ns
Goldf	12				0	0.72**
Linne	25				0	–0.35 ns

Table 2. Comparison of wing length as a measure of body size (through its prediction of variance in body mass) with principal component 1 (PC1) derived from a PCA of wing, tarsus, tail and head lengths. For each of 26 species (for abbreviations see Appendix 2) for which at least 12 individuals were measured, a stepwise multiple regression analysis was performed with mass as the dependent variable and either wing length or PC1 plus fat and muscle scores, as predictors. The table shows independent r^2 values for significant predictors.This table should be considered together with Table 1.

Species	n	*Wing*	*Fat*	*Muscle*	*Total*	*PC1*	*Fat*	*Muscle*	*Total*	r *F&M*
Skyla	15	30.8	27.2	15.4	73.3	24.6	39.9	10.4	74.9	0.49*
Meapi	159	16.1	25.8		41.9	15.4	25.8	1.7	42.9	0.20*
Rocpi	139	43.7	5.0		48.6	48.3	4.2		52.5	0.19*
Piewa	15				0				0	–0.26 ns
Dunno	33	9.0	47.5		56.6	12.0	47.5		59.6	0.32*
Robin	83		30.9	10.2	41.1	6.7	30.9	10.2	47.8	0.29***
Redst	21			23.3	23.3			23.3	23.3	0.18 ns
Wheat	219	7.4	47.0		54.3	7.8	47.0		54.8	0.28***
Blabi	514	10.2	26.6	0.9	37.7	14.9	26.6	0.9	42.4	0.30***
Sonth	65		35.0		35.0		35.0		35.0	0.39**
Redwi	299	18.3	32.7		51.0	37.9	22.3		60.2	0.40***
Leswh	24		40.6		40.6		40.6		40.6	0.57***
White	19		32.9		32.9	47.7	19.9		67.6	0.15 ns
Garwa	45		54.7		54.7	6.8	54.7		61.5	0.38**
Blaca	92	5.5	56.6	1.7	63.8	7.0	56.6	2.5	66.1	0.44***
Chiff	90	35.1	12.8	6.8	54.7	37.3	18.4	7.5	63.1	0.12 ns
Wilwa	82	31.2	18.2		49.4	38.5	17.3		55.8	0.08 ns
Goldc	59		23.4		23.4	6.4	23.4		29.8	0.46***
Spofl	24	19.8		33.4	53.3	17.5		17.1	34.7	0.32***
Piefl	14				0				0	`–0.11 ns
Chaff	25		46.6		46.6		46.6		46.6	0.39*
Bramb	109	15.1	30.1		45.2	14.8	30.1		44.9	0.33***
Twite	117	5.6	6.0		11.7	4.9	6.0		11.0	0.25**
Redpo	13				0				0	0.39 ns
Cross	29	18.3	34.6		52.9	9.3	34.6		43.9	0.27 ns
Reebu	13	43.2		37.8	80.9	61.5	7.0	18.5	86.9	0.43*

greatly) and muscle are not correlated.[18]

Table 2 also compares the values of wing length and PC1 (derived from length of wing, maximum tarsus, tail and head) as predictors of mass in Group 3 samples after correcting for fat and muscle scores. In 12 (46%) out of 26 species some improvement in r^2 explained by size was found by using PC1. However, in six species wing length explained more variance in mass than PC1 and the overall mean improvement by using PC1 was only 6.4%. On the basis of this, wing length would appear to be a reasonable alternative to PC1 if it is not possible to take more measurements.

Comparison of body dimensions to predict mass

The analyses above showed that PC1 based on four dimensions was generally better than wing length as a predictor of mass, and therefore size, but that the improvement was generally small. To test which of the four was the best single measure of size, we carried out stepwise multiple regression analyses on data from each species with wing, maximum tarsus, tail, head, fat and muscle as predictors and mass as dependent. From this we determined: (a) which was the best size predictor after controlling for fat and muscle condition, (b) which measures contributed to the significant set of predictors, (c) what was the next best predictor if the best was dropped from the analysis (which indicates whether, if measurements were highly correlated, an alternative might be suitable) and (d) the total r^2 explained by all six predictors, for comparison with Table 2. Table 3 summarizes the results and

Table 3. Significant predictor sets to predict mass. ♦, Best size predictor (included for fat/muscle if these are best overall predictors); *, included in the significant predictor set; +, included when best size predictor is dropped from predictor set. For abbreviations see Appendix 2.

Species	n	*Wing*	*Tarsus*	*Tail*	*Head*	*Fat*	*Muscle*	*Best* r^2
Skyla	15	♦	+			*+	*	73.3
Meapi	159	♦	*+	+		♦+	*+	45.6
Rocpi	139	♦	*+	+	+	*+		55.2
Piewa	15							0
Dunno	33	♦		+		♦+		56.6
Robin	83		♦	*+		♦+	*+	61.1
Redst	21				♦	*	*+	63.3
Wheat	219	♦	+	+	+	♦+		54.3
Blabi	514	♦	*+	+	*+	♦+	*+	43.4
Sonth	65	+			*	♦+	+	39.8
Redwi	299	*+	♦	+	*+	*+		61.5
Leswh	24			♦		♦+		54.1
White	19			♦	*+	*+		84.2
Garwa	45				♦	♦+		59.6
Blaca	92	♦		+	*+	♦+	*+	66.6
Chiff	90	♦	+	*+	*+	*+	*+	62.6
Wilwa	82	♦		*+	*+	*+		57.2
Goldc	59		♦			♦+		28.7
Spofl	24	♦		+	*+		♦+	61.7
Piefl	14							0.0
Chaff	25					♦		46.5
Bramb	109	♦		+		♦+		45.2
Twite	117	♦				♦+		11.7
Redpo	13							0.0
Cross	29	♦		*		♦+	*	73.9
Reebu	13	+	+		♦		+	56.1
***N* best**		13	3	2	3	14	1	
Best set		14	6	6	11	21	9	
***N* all**		16	10	15	13	21	11	

Appendix 1 gives the independent r^2 values for each predictor.

As was found in the larger data set, fat was a strong predictor of mass in most species, the best predictor in 14 species out of 26 (54%) and in the best set of predictors in 21 (81%). Muscle was best predictor just once, but included as a significant predictor in nine species (35%). Of the four size predictors, wing length was consistently better than the others. It was best predictor in 13 species (50%) and included in the significant set in 14 (54%). Tarsus and head lengths were best in just three species each and tail length in just two. However, although tail was included six times in the significant set, this was doubled if the best predictor (wing length) was dropped. This is because wing and tail lengths were generally highly correlated within species, presumably because of their close functional relationship[21] (mean $r_{\text{wing.tail}}$ across 26 species = 0.66 ± 0.15 sd; c.f. $r_{\text{wing.tarsus}}$ = 0.29 ± 0.22; $r_{\text{wing.head}}$ = 0.31 ± 0.18; $r_{\text{tarsus.tail}}$ = 0.16 ± 0.26; $r_{\text{tarsus.head}}$ = 0.23 ± 0.20; $r_{\text{head.tail}}$ = 0.25 ± 0.22).

Comparing total r^2 values in Tables 2 (right-hand column) and 3 reveals that, in some species, considerably more variance in mass is explained by entering the four measures separately than by combining them as a principal component. This suggests that there might be variance in size within individual dimensions which is either due to overall body size and independent of the other linear dimensions or is correlated with part of the unexplained variance in mass. This might be a component of condition if, for example, birds of a species with longer bills were heavier because of some aspect of condition not reflected in fat or muscle (e.g. fuller gut).

Reproducibility of measurements

One or more repeat measurements were submitted for 1422 birds; an analysis of reproducibility of wing measurements from this whole data set has already been published.[14] In this paper, analysis is limited to those birds for which repeat data of all measurements were obtained and to species for which at least two individuals were measured. This means that, although in a few species sample sizes were small, the reproducibility statistics for different measurements are directly comparable. Because a slightly different set of measurements was requested in 1989, we have analysed data from this set (Set A, 233 birds of 24 species) separately from the rest (Set B, 68 birds of 15 species) (Appendix 2).

Table 4 gives mean reproducibility statistics

Table 4. Reproducibility of measurements. Measures of bias, disagreement and imprecision (see text) are given, averaged (± 1 sd) across 30 species of passerine ranging in size from Goldcrest to Mistle Thrush. Note that as imprecision is a variance (error mean square) its units are mm^2. Statistics were derived from two data sets differing in the variables measured. Set A (n = 24 species), collected in 1989, contains wing, maximum tarsus, minimum tarsus, head, fat and muscle. Set B (n = 15 species), collected in 1990 and 1991, contains tail but not minimum tarsus.

Metric	*Bias (mm)*	*Disagreement (mm)*	*Imprecision (mm^2)*
Set A			
Wing	–0.34 ± 0.347	0.44 ± 0.306	0.22 ± 0.258
Maximum tarsus	0.36 ± 0.602	0.68 ± 0.430	0.56 ± 0.568
Minimum tarsus	0.66 ± 0.503	0.89 ± 0.570	0.73 ± 1.128
Head	–0.04 ± 0.300	0.41 ± 0.222	0.16 ± 0.201
Fat[a]	–0.30 ± 0.374	0.37 ± 0.354	0.18 ± 0.140
Muscle[a]	–0.01 ± 0.135	0.09 ± 0.124	0.06 ± 0.097
Set B			
Wing	–0.41 ± 0.527	0.76 ± 0.446	0.74 ± 0.770
Maximum tarsus	0.35 ± 0.518	0.52 ± 0.506	0.49 ± 1.104
Tail	–0.53 ± 0.804	1.08 ± 0.528	1.09 ± 0.910
Head	0.15 ± 0.580	0.65 ± 0.419	0.46 ± 0.650
Fat[a]	–0.06 ± 0.334	0.34 ± 0.209	0.19 ± 0.131
Muscle[a]	0.13 ± 0.335	0.26 ± 0.250	0.11 ± 0.107

[a]No dimensions.

Table 5. Reproducibility measures as in Table 4 but scaled as percentages of the species' standard deviation for bias and disagreement, or variance for imprecision (as this is itself a variance measure; see Table 4) for the measurement concerned.

Metric	*Bias (%)*	*Disagreement (%)*	*Imprecision (%)*
Set A (24 species)			
Wing length	–10.4 ± 11.26	14.0 ± 8.84	2.7 ± 2.19
Maximum tarsus	53.3 ± 144.40	69.3 ± 139.10	31.1 ± 65.10
Minimum tarsus	90.8 ± 76.10	134.0 ± 124.60	281.0 ± 867.00
Head length	28.3 ± 159.30	75.3 ± 147.80	20.1 ± 21.06
Fat score	–26.6 ± 34.20	31.8 ± 32.25	15.8 ± 22.75
Muscle score	–1.8 ± 25.70	16.8 ± 22.82	15.7 ± 21.88
Set B (15 species)			
Wing length	–17.9 ± 24.83	31.3 ± 20.31	12.5 ± 14.16
Maximum tarsus	39.0 ± 56.70	69.4 ± 98.70	136.0 ± 380.10
Tail length	–17.7 ± 29.09	37.6 ± 20.45	13.0 ± 11.13
Head length	–10.5 ± 62.00	63.3 ± 38.71	41.4 ± 46.70
Fat score	–6.1 ± 25.49	26.2 ± 18.29	11.9 ± 12.11
Muscle score	26.3 ± 63.60	48.2 ± 49.60	37.1 ± 35.54

for each measurement in the two sets. In Table 5, these statistics are averaged after scaling within species relative to the standard deviation or, in the case of imprecision, the variance for the measure concerned (see Methods). Harper[15] suggests that reproducibility measures (which he called repeatability) greater than 70% are high, and above 90% very high. In these terms, for linear measures, only the reproducibility for wing length could be described as very high in any case, whereas in some cases reproducibility was very poor. Thus wing length was consistently better than other measures and up to two orders of magnitude better than tarsus measurements.

The pre-eminence of wing reproducibility may partly reflect the familiarity of participants with the different measures (all had been previously trained to take wing measurements whereas few were experienced at measuring tarsus or head). Another important consideration here concerns the natural variation in size of parts being measured.

Table 6. Mean range for each measurement across all species where at least 12 individuals were measured.

Measurement	*Mean range (mm) ± 1 sd*	n *species*
Wing length	14.9 ± 11.94	27
Maximum tarsus	3.6 ± 3.23	27
Minimum tarsus	1.9 ± 1.32	19
Tail	4.8 ± 4.65	21
Head	3.6 ± 5.08	26

Table 6 shows the mean range (in mm) of each measurement across a range of species. Wing length is consistently more variable than other measures. Hence a small error in measurement, due to the scale of the observer relative to that of the part measured, will be relatively less important for wing than for other parts.

We found little difference in reproducibility between the two methods of measuring tarsus and a paired *t*-test across the 24 species of Set A found no significant difference ($t_{24\text{paired}} = 0.76$ ns). Reproducibility of fat and muscle, which were new to all observers and which naturally require more 'judgement' than linear measurements, were very much poorer than linear measures in absolute reproducibility (Table 4). This was also to be expected, especially for muscle, because of the comparatively small number of classes (six for fat, three for muscle) available for observation. Relative to the natural variation of the measures, however, they were rather better than linear measures, with 'high' or even 'very high' reproducibility in Harper's terminology.[15] With training we suggest that yet further improvement would be achieved. Furthermore, the fact that fat score was consistently a strong predictor of mass indicates that, even with a degree of error, reliable information is contained in the data.

CONCLUSIONS

Correlation analyses of the type presented here may be strongly influenced by lack of variance in any predictor, perhaps due to inadequate or biased sampling, so little weight should be attached to the precise value of individual correlations or reproducibility statistics (some of which were based on very small samples). Rather, it is the pattern indicated by assessing data across many species that is important. Thus, taking a representative set of passerines (with a bias toward migrating birds), although wing length may not be best for every species, it is generally best.

Our analyses show, therefore, that where only one measurement can be taken, it should be wing length because it is consistently the best predictor of estimated lean mass (size) and is the most reproducible. Wing length tends to be more variable in absolute terms (rather than scaled to the mean) than other dimensions within species, which reduces the relative importance of measurement error. The failure of other measures to give better prediction of size than wing length cannot be attributed to their slightly poorer reproducibility because fat score, which is very much less reproducible, is a better predictor of mass even than wing length. Where measurement of size is particularly critical and logistical considerations allow, PC1 based on several measures may be preferred[9] although our data indicate that it generally adds rather little (sometimes nothing). Wing length may be a poor measure of size in the genus *Sylvia*; we did not find this to be so in other sylviid genera. Further investigation is required to discover whether this is really a characteristic of the genus and what might be preferred measures of size for these species.

Our data cast some doubt on the use of tarsus length, which in recent years has increasingly been promoted as a measure of size in passerines.[9] There are several reasons for this. The main one is that unlike wing length which, in many species, increases with successive moults, tarsus length is full-grown at fledging and remains constant thereafter. For example, Garnett[22] studied carcasses of 15-day-old Great Tit nestlings. He found that while, by this age, mass and tarsus were fully developed, the wings and tail were not, and so gave a poor reflection of overall size. Thus while tarsus length may be the best size measure for nestlings, the case for its general use as a measure of body size for full-grown birds is not proven.

Finally it is interesting to consider why wing length might generally be the best measure of size. A clue to this may be in the strong correlations we found between wing and tail, since a strong relation between measures of the flight surfaces and the payload (i.e. wing, tail and mass) could be under strong selection for optimal flight performance.

ACKNOWLEDGEMENTS

We are indebted to the Bird Observatory staff who contributed the data presented here, in particular: P. Jenks at Bardsey Island B.O.; D. Walker at Dungeness B.O.; A. Sapsford, I. Fisher and P. Naylor at Calf of Man B.O.; P. Harvey, P. Howlett, A. Silcox, C. Donald and R. Proctor at Fair Isle B.O.; A. Duncan, I. Fisher and K. Woodbridge at North Ronaldsay B.O.; M. Rogers at Portland B.O.; I. Hunter at Sandwich Bay B.O.; and D. Bowes and I. Crowther at Spurn B.O. We are also grateful to 30 visiting ringers (known to us only by their initials) who took part in the reproducibility trials. We are grateful to the British Ornithologists' Union for the award of a small projects grant to cover the cost of data preparation. Finally we thank Dr David Harper, Professor Jim Rising and Dr Adrian Thomas for their valuable comments on an earlier draft.

REFERENCES

1. Harvey, P.H. & Pagel, M.D. (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
2. Peters, R.H. (1983) *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge.
3. Newton, I. (1979) *Population Ecology of Raptors*. T & A.D. Poyser, Berkhamstead.
4. Schmidt-Nielsen, K. (1984) *Scaling. Why is Animal Size so Important*. Cambridge University Press, Cambridge.
5. Berthold, P. (1993) *Bird Migration, a General Survey*. Oxford University Press, Oxford.
6. Berthold, P. (1996) *Control of Bird Migration*. Chapman & Hall, London.

7. Perrins, C.M. (1964) Weight. In *A New Dictionary of Birds* (ed. A. Landsborough Thomson), pp: 883–884. Nelson, London.
8. Connell, C.E., Odum, E.P. & Kale, H. (1960) Fat-free weights of birds. *Auk,* **77**, 1–9.
9. Rising, J.D. & Somers, K.M. (1989) The measurement of overall body size in birds. *Auk,* **106,** 666–674.
10. Spencer, R. (1965) *The Ringer's Manual*, 1st edn. BTO, Tring.
11. Svensson, L. (1970) *Identification Guide to European Passerines*, 1st edn. Naturhistoriska, Stockholm.
12. Flegg, J. (1973) A study of treecreepers. *Bird Study,* **20,** 287–302.
13. Berthold, P. & Friedrich, W. (1979) Die Federlänge: ein neues nützliches Flügelmass. *Vogelwarte,* **30**,11–21.
14. Gosler, A.G., Greenwood, J.J.D., Baker, J.K. & King, J.R. (1995) A comparison of wing length and primary length as size measures for small passerines: a report to the British Ringing Committee. *Ring. Migr.* **16,** 65–78.
15. Harper, D.G.C. (1994) Some comments on the repeatability of measurements. *Ring. Migr*. **15,** 84–90.
16. Spencer, R. (1984) *The Ringer's Manual*, 3rd edn. BTO, Tring.
17. Gosler, A.G. (1996) Environmental and social determinants of winter fat storage in the Great Tit *Parus major*. *J. Anim. Ecol*., **65,** 1–17.
18. Gosler, A.G. (1991) On the use of greater covert and pectoral muscle as measures of condition with data for the Great Tit *Parus major*. *Bird Study,* **38,** 1–9.
19. Svensson, L. (1992) *Identification Guide to European Passerines*, 4th edn. Svensson, Stockholm.
20. Biebach, H. (1996) Energetics of winter and migratory fattening. In *Avian Energetics and Nutritional Ecology*, (ed. C. Carey), pp. 280-323. Chapman & Hall, London.
21. Thomas, A.L.R. (1996) Why do birds have tails? The tail as a drag reducing flap, and trim control. *J. Theor. Biol*., **183,** 247–253.
22. Garnett, M.C. (1981) Body size, its heritability and influence on juvenile survival among Great Tits, *Parus major*. *Ibis,* **123,** 31–41.

APPENDIX 1

Significant sets to predict mass. Data used for Table 1.

Species	n	*Wing*	*Tarsus*	*Tail*	*Head*	*Fat*	*Muscle*	*Best* r^2
Skyla	15	30.8				27.2	15.4	73.3
Meapi	159	16.1	2.3			25.8	1.4	45.6
Rocpi	139	43.7	7.2			4.4		55.2
Piewa	15							0.0
Dunno	33	9				47.5		56.6
Robin	83		21.4	2.3		30.9	6.5	61.1
Redst	21				25.2	14.7	23.3	63.3
Wheat	219	7.4				47		54.3
Blabi	514	10.2	1.8		4.1	26.6	0.8	43.4
Sonth	65				4.7	34		39.8
Redwi	299	2.4	34.6		1.2	23.3		61.5
Leswh	24			13.5		40.6		54.1
White	19			37.5	21.9	24.7		84.2
Garwa	45				5	54.7		59.6
Blaca	92	5.5			2.3	56.6	2.2	66.6
Chiff	90	35.1		2.6	5.4	12.8	6.8	62.6
Wilwa	82	31.2		3.9	4	18.2		57.2
Goldc	59		5.3			23.4		28.7
Spofl	24	19.8			8.5		33.4	61.7
Piefl	14							0.0
Chaff	25					46.5		46.5
Bramb	109	14.1				30.1		45.2
Twite	117	5.6				6		11.7
Redpo	13							0.0
Cross	29	18.3		11		34.6	10	73.9
Reebu	13				56.1			56.1

APPENDIX 2

Scientific names and abbreviations for species used in the text. 'Repeats' is the set (A or B) used in reproducibility analyses: A included minimum tarsus but not tail, B included tail instead of minimum tarsus.

Species	*Abbrev.*	*Scientific name*	*Repeats*
Skylark	Skyla	*Alauda arvensis*	
Sand martin	Sanma	*Riparia riparia*	A
Swallow	Swall	*Hirundo rustica*	A
Tree Pipit	Trepi	*Anthus trivialis*	B
Meadow Pipit	Meapi	*Anthus pratensis*	B
Rock Pipit	Rocpi	*Anthus petrosus*	
Yellow Wagtail	Yelwa	*Motacilla flava*	A
Pied Wagtail	Piewa	*Motacilla alba*	
Wren	Wren	*Troglodytes troglodytes*	A
Dunnock	Dunno	*Prunella modularis*	A
Robin	Robin	*Erithacus rubecula*	A, B
Black Redstart	Blare	*Phoenicurus ochruros*	
Redstart	Redst	*Phoenicurus phoenicurus*	
Whinchat	Whinc	*Saxicola rubetra*	A
Wheatear	Wheat	*Oenanthe oenanthe*	B
Ring Ouzel	Rinou	*Turdus torquatus*	
Blackbird	Blabi	*Turdus merula*	A, B
Fieldfare	Field	*Turdus pilaris*	
Song Thrush	Sonth	*Turdus philomelos*	A, B
Redwing	Redwi	*Turdus iliacus*	
Mistle Thrush	Misth	*Turdus viscivorus*	A
Reed Warbler	Reewa	*Acrocephalus scirpaceus*	A
Sedge Warbler	Sedwa	*Acrocephalus schoenobaenus*	A
Barred Warbler	Barwa	*Sylvia nisoria*	
Lesser Whitethroat	Leswh	*Sylvia curruca*	A, B
Whitethroat	White	*Sylvia communis*	A, B
Garden Warbler	Garwa	*Sylvia borin*	A, B
Blackcap	Blaca	*Sylcia atricapilla*	A, B
Chiffchaff	Chiff	*Phylloscopus collybita*	A, B
Willow Warbler	Wilwa	*Phylloscopus trochilus*	A, B
Goldcrest	Goldc	*Regulus regulus*	B
Spotted Flycatcher	Spofl	*Muscicapa striata*	B
Pied Flycatcher	Piefl	*Ficedula hypoleuca*	B
Great Tit	Greti	*Parus major*	
Red-backed Shrike	Rebsh	*Lanius collurio*	
Starling	Starl	*Sturnus vulgaris*	
Chaffinch	Chaff	*Fringilla coelebs*	
Brambling	Bramb	*Fringilla montifringilla*	
Goldfinch	Goldf	*Carduelis carduelis*	A
Linnet	Linne	*Carduelis cannabina*	A
Twite	Twite	*Carduelis flavirostris*	
Redpoll	Redpo	*Carduelis flammea*	A
Crossbill	Cross	*Loxia curvirostra*	
Bullfinch	Bullf	*Pyrrhula pyrrhula*	A
Corn Bunting	Corbu	*Miliaria calandra*	A
Reed Bunting	Reebu	*Emberiza schoeniclus*	A

(MS received 27 February 1997; revised MS accepted 20 June 1997)