Weighing with spring balances: Biases and errors

Charles T. Collins and Jonathan L. Atwood

In the course of their field work most banders accumulate large amounts of data on the birds they handle. One of the most frequently recorded parameters is body weight. While workers at permanent banding stations may enjoy the luxury of sophisticated equipment such as triple-beam balances (Clench and Leberman 1978), or even electronic balances as at Point Reyes Bird Observatory, most banders operate under field conditions where they must rely on spring balances for obtaining body weight data (Collins 1972; Collins and Bradley 1971; Atwood 1979). Even though such balances appear to be relatively accurate under ideal testing condition, possible differences in individual investigators' accuracy in reading their scales raise questions about the reliability of data obtained through such methods.

In addition — since recorded values usually include a final digit which, due to the balances' calibration interval, must be visually estimated — the possibility of errors introduced by investigator's subconscious biases must be considered. Such potential sources of inaccuracy are of particular concern when detailed analyses are made using such data (Collins and Bradley 1971; Collins and Atwood in prep.) or when data collected by several participants in a cooperative banding project are to be compared (Mewaldt 1975, 1976; Mewaldt and King 1978). We suspect that many banders are "hoarding" large amounts of body weight data in their field notes, at least partly as a result of having little confidence in its accuracy. However, the extent of investigator biases and weighing errors does not appear to have been fully considered and is the subject of our analysis.

In this paper we consider (1) whether biases are evident in investigators' estimates of the final digits of body weight recorded under field conditions, (2) whether investigator biases and/or weighing errors occur in spring-balance weight data obtained under experimental conditions, and (3) whether inaccuracies encountered under experimental test conditions result in any significant alteration of the final calculated results.

In Table 1 we have summarized a variety of our personally obtained weight data as well as some collected by various individuals during the 1970 and 1971 cooperative banding project at Morongo Valley, California (Collins and Bradley 1971 and unpublished). The assumption for statistical testing was that in the final estimated digits of the recorded weights all values from 0 to 9 should occur with equal frequency (10%). Although we made no effort to select samples which showed evident biases, 8 of the 9 deviated significantly from the expected random frequency distribution ($P<0.01$, Chi-square test), with the 9th sample (Wilson's Warbler, Morongo Valley 1971) also deviating from expected at a slightly lower significance level ($P<0.05$).

Without control body weights obtained on more accurate equipment, it is impossible to conclude with certainty that all of these samples indicate the presence of estimation biases, since the true weight values themselves may actually have deviated from a random frequency distribution. However, close examination of the data makes investigator bias highly suspect as the source of these non-random distributions. Each of the 9 samples shows some tendency toward over-use of 0 as the final recorded digit; that is, there was a bias for recording body weights at the even or whole-gram level. This bias was most evident in the Morongo Valley Swainson's Thrush samples and in all 3 samples handled by Atwood. Additionally, all 3 Morongo Valley samples estimated 5, and Collins — in the Least Tern Studies — seemed to estimate 4, more frequently than expected. These and some of the other departures from randomness may be due to an unconscious effort to avoid rounding off values. The result is an over-utilization of digits on one or both sides of the scale marks of the balance (Table 1.).

To examine further the presence and potential influence of these apparent investigator biases associated with spring-balance weighing, we performed a simple experiment in which 11 participants each weighed 75 bags of sand (approximate weight 40 grams). These weights were obtained using a 50-gram Pesola balance calibrated in 0.5-gram intervals; all weights were estimated to the nearest 0.1 gram. The sample bags were also weighed on an electronic balance to provide a known standard for comparison. Participants in the experiment included 3 individuals (CC, JA, DB) who had several years of experience weighing birds with

Table 1. Probable investigator bias in spring-balance weighing data

Species (Locality & date)	Bander	Occurence of final estimated digits											X^2
		N	0	1	2	3	4	5	6	7	8	9	
Wilson's Warbler[1] (Morongo Valley, 1971)	Cooperative	540	71	47	46	52	54	66	37	45	66	56	19.78*
Swainson's Thrush[1] (Morongo Valley, 1971)	Cooperative	384	87	25	38	24	31	67	16	23	42	31	115.55**
Swainson's Thrush[1] (Morongo Valley, 1972)	Cooperative	206	74	16	20	10	16	28	10	9	12	11	168.65**
Least Tern[2] (L.A. Co., Orange Co., 1979)	Collins	402	50	22	30	36	49	60	38	31	47	39	28.74**
Least Tern[2] (L.A. Co., Orange Co., 1980)	Collins	395	46	13	29	43	60	44	33	47	48	32	38.83**
House Sparrow[2] (Long Beach, 1978-79)	Collins	211	29	13	22	27	40	24	15	12	14	15	34.91**
Orange-crowned Warbler[2] (Goleta, 1971-72)	Atwood	273	63	25	32	25	20	35	18	16	15	24	65.78**
Wilson's Warbler[2] (Goleta, 1971-72)	Atwood	170	29	24	10	20	15	25	8	7	13	19	30.58**
Scrub Jay[3] (Santa Cruz Island, 1975-77)	Atwood	523	110	58	53	62	35	60	35	31	28	51	98.65**

[1] Weighed on 30-gram Pesola balance calibrated in 1.0-gram intervals; weights recorded to nearest 0.1-grams.
[2] Weighed on 50-gram Pesola balance calibrated in 0.5-gram intervals; weights recorded to nearest 0.1-grams.
[3] Weighed on 300-gram Pesola balance calibrated in 2.0-gram intervals; weights recorded to nearest 1.0-grams.
Significant deviation from random frequency distribution: * $P<0.05$, ** $P<0.01$

Table 2. Experimental analysis of investigator biases in spring-balance weighing

Estimated final digit	Numbers of bag weights actually ending in indicated final digit[1]	Number of times estimated final digit was recorded by each participant										
		DB	JA	CC	TP	MA	BB	CB	JV	JR	TH	KK
0	8	4	10	11	15	11	10	13	8	23	13	8
1	11	10	7	6	8	9	11	7	7	1	14	9
2	9	11	11	9	5	10	6	10	12	7	3	9
3	5	8	3	9	7	5	4	4	5	9	3	6
4	3	3	6	6	3	3	4	4	5	5	4	6
5	10	6	8	9	14	12	14	13	9	9	12	8
6	6	7	7	5	5	4	7	3	6	5	5	3
7	6	5	6	3	4	6	4	6	5	1	4	7
8	6	13	9	7	7	7	3	8	5	2	7	9
9	11	8	8	9	7	8	12	7	13	3	10	8
X^2		15.25	7.45	10.36	13.55	3.62	5.93	9.74	4.58	49.64*	10.54	5.71

[1] Based on values obtained by means of an electronic balance.
* Significant difference, $P<0.01$

Table 3. Effects of weighing inaccuracies on final calculated results

No.	Range of values[1] grams	Estimation level grams	N	Actual values Mean±S.D.	Inaccurate values Mean±S.D.	Source of inaccuracy
1	7.0 - 10.0	0.1	75	8.491±0.733	8.495±0.713	Estimation bias. All 0.1 and 0.9 values recorded as 1.0; all 0.4 and 0.6 values recorded as 0.5
2	7.0 - 10.0	0.1	75	8.491±0.733	8.397±0.789	Careless error. 10% of actual values recorded one gram off (low).
3	7.0 - 10.0	1.0	75	8.491±0.733	8.920±1.080*	Balance and object poorly matched (estimation required at level of 10-14% of object's weight). Estimation bias - 75% of actual values of 9 recorded as 10.
4	110 - 130	1.0	75	117.91±5.722	118.01±5.840	Estimation bias. All final digit values of 4 recorded as 5.
5	110 - 130	1.0	75	117.91±5.722	117.80±6.516	Estimation bias. All final digit values of 1 or 9 recorded as 0.

[1] Derived from random number table.
* Inaccurate mean significantly different from actual mean ($P<0.05$)

spring balances and 8 members of the 1980 Field Ornithology class (CSULB) with limited or no prior experience with such balances. Test conditions did not exactly mirror those often encountered in the field, since (1) there was no wind (weights were taken indoors), (2) weights were measured from a suspended rather than hand-held balance, (3) no time constraints were placed on the participants, and (4) the bags of sand were substantially less active when being weighed than wild birds in the field.

Table 2 summarizes the results of the spring-balance weighing experiment. Only 1 (JR) of the 11 participants obtained results which differed significantly ($P<0.001$) from the actual frequency distribution; this individual displayed a strong bias for over-using 0 and under-using 1 and 9 as the final recorded digit. The fact that more individuals did not display estimation biases may be related to the absence of time constraints, which probably encouraged increased accuracy.

Interestingly, we also found 14 errors (1.7% of the entire sample) in recorded whole-gram weights. Such errors presumably occurred when observers misread the gram scale while concentrating on estimation of the 0.1-gram interval. Two of these errors were greater than 1-gram: KK and JR each recorded 5-gram errors. Only 4 of the 11 participants (DB, MA, BB, and TH) did not make any whole-gram errors.

It is evident from these results that investigator biases and careless errors occur regularly in the process of spring-balance weighing. The question remains: how much effect do such inaccuracies have on the final results calculated from raw data?

Despite varying degrees of accuracy in the experimental test data, mean values derived from the results of each participant were highly comparable, with all of the spring balance means falling within the 95% confidence interval of the actual mean obtained from the electronic balance (40.97-grams ± 0.966). Importantly, this even includes the data reported by JR, who not only displayed a significant bias toward recording whole, rather than decimal, values, but who also made 4 careless errors totalling 8-grams. This somewhat surprising result is further elaborated in Table 3, which is based on a hypothetical set of random weight values which have been intentionally biased or carelessly recorded. In these results, mean values based on various types of inaccurate data did not differ at the 0.05 level of significance from means based on accurate values; this was true not only when estimation biases were introduced at the 0.1-gram level but also (for weights ranging from 110-130) at the 1.0-gram level. An exception to this general "resistance" of calculated results to minor errors in the raw data is indicated in sample 3 (Table 3); in this case, when investigator bias was introduced in a situation where the interval between balance calibrations represented a major percentage of the weight of the objects being sampled, the difference between the actual and the inaccurate means was significant at the 0.05 level.

Therefore, it appears that under most circumstances where adequate samples are available the inaccuracies associated with spring-balance weighing techniques do not significantly alter the final conclusions; in general, spring-balances provide surprisingly accurate results. Accordingly, it seems reasonable to place greater reliance on weight data obtained under field conditions, as well as on analyses based on such data.

Finally, although minor investigator biases and errors seem to have little impact on the reliability of final weight results, we would encourage banders nonetheless to evaluate carefully their spring-balance weighing techniques along the following lines:

a) Determine whether there are some integers which occur with unexpected frequency as estimated digits in your raw data. Mere awareness that you tend to bias your weight estimations in a certain way may do much to correct the problem in the future.

b) Be cautious of becoming subconsciously "fixed" on certain weight values. This possibility seems especially likely when large numbers of individuals of a single species are being weighed during a relatively short time period; if a certain weight value occurs repeatedly within a small sample, it would seem easy to begin subconsciously thinking of this weight value as the "usual" weight for "most" individuals being handled. There may even be some value in breaking up the weighing of very common species with less common ones — if for no other reason than mentally escaping the tedium which often sets in when banding large numbers of the same species.

c) Attempt to suspend spring-balances from a stationary object rather than holding them by hand. Movement of the bird itself makes reading such balances difficult enough, without the bander contributing his own arm motion. Obviously, wind disturbance should be avoided when possible, and methods should be used which adequately restrain the bird's attempts to escape.

d) Be aware that it is easy to make careless errors in recording the whole-gram or — with larger balances — the 10-gram values.

e) Make sure that the required level of estimation on the balances which you use "matches" the weights of the birds being sampled. We realize that the cost of spring-balances is no longer minimal and that most banders may only be able to afford one or two sizes. It is probably best to pick a balance which will accurately handle your most common species, since you may never accumulate enough raw data to analyze in detail the body weights of less common species.

f) Last of all, publish some of the body-weight data which you have carefully hidden in your field notes! Such data are very easy to analyze, and although they may be of little value to your particular research interests, they may be very helpful to other investigators.

Summary

Body weight data obtained under field conditions are subject to an array of estimation biases. However, under laboratory conditions mean weights obtained with spring-balances are not significantly different from the ones obtained electronically. In fact, with adequate sample sizes spring-balance-obtained mean weights appear quite "resistant" to an array of simulated errors and biases.

Acknowledgments

We are grateful to the many banders who helped generate the Morongo Valley weight data utilized here and to Martha H. Balph for helpful comments on the manuscript. Dave Bontrager and the members of the Field Ornithology class helped to provide the experimental weight data, particularly JR whose data made this analysis more interesting; but now you know who shot JR!

Literature cited

Atwood, J.L. 1979. Body weights of the Santa Cruz Island Scrub Jay. *N. Am. Bird Bander* 4:148-153.

Clench, M.H. and R.C. Leberman. 1978. Weights of 151 species of Pennsylvania birds analyzed by month, age and sex. *Bull. Carnegie Mus. Nat. Hist.* 5:1-85.

Collins, C.T. and R.A. Bradley. 1971. Analysis of body weights of spring migrants in southern California. *West. Bird Bander* 46:38-40 and 46:48-51.

Collins, C.T. 1972. Weights of some birds of North-central Venezuela. *Bull. Brit. Ornith. Club* 92:151-153.

Mewaldt, L.R. 1975. White-crowned Sparrow project 1974-75. *West. Bird Bander* 50:50-51.

Mewaldt, L.R. 1976. White-crowned Sparrow cooperative project, present and future. *N. Am. Bird Bander* 1:87.

Mewaldt, L.R. and J.R. King. 1978. Latitudinal variation in prenuptial molt in wintering Gambel's White-crowned Sparrows. *N. Am. Bird Bander* 3:138-143.

Department of Biology, California State University, Long Beach, CA 90840 and Department of Biology, University of California, Los Angeles, CA 90024.