A STANDARDIZED PROTOCOL FOR RULER-BASED MEASUREMENT OF WING LENGTH IN MONARCH BUTTERFLIES, *DANAUS PLEXIPPUS* L. (NYMPHALIDAE, DANAINAE)

Tonya Van Hook [1,5], Ernest H. Williams [2], Lincoln P. Brower [1], Susan Borkin [3], and Julie Hein [4]

[1]Biology Department, Sweet Briar College, Sweet Briar, VA 24595; [2]Biology Department, Hamilton College, Clinton, NY 13323; [3]Invertebrate Zoology, Milwaukee Public Museum, Milwaukee, WI 53233; [4]5904 Cedar Creek Place, Sturgeon Bay WI 54235

[5]Corresponding author, Email: tonyavanhook@yahoo.com, Tel: (423) 914-0842

Abstract - Standardized measurements using well-defined landmarks are the most effective means to reduce measurement error. We describe such a protocol for monarch forewings based on single measurements with a ruler to the nearest 1.0 mm. Analysis of this protocol showed that it provides excellent intra-observer repeatability, excellent to substantial inter-observer repeatability, and similar wing length estimates as those of calipers at 0.1 mm, as long as sample sizes are ≥ 30. In addition, our study showed that males and females differ in wing length; different observers differ in their measurements and in their measurement error; and wings shrink slightly when dried. We make these recommendations for study of monarch wing lengths: 1) males and females should be analyzed separately; 2) live butterflies should be measured after cooling and dead butterflies should be measured before they are dried; 3) measurements should be restricted to the right forewing; 4) the standard protocol should be practiced and calibrated until measurements are repeatable within and among measurers; 5) the samples should be mixed among all observers when possible to mitigate relative biases; and 6) names, handedness, measurement error, and archived raw data should be reported. Widespread adoption of this protocol will increase the comparability of wing length data from various investigators. Similarly based standardization of measurement would benefit wing measurement of all Lepidoptera.

Key words: standardized measurement protocol, butterfly wings, body size, methodology

INTRODUCTION

Forewing length is the most commonly used measure of body size in Lepidoptera (Miller 1977, 1991), and wing length is often used in studies of monarch butterflies (Table 1). Wing length is correlated with wing width and area (Altizer & Davis 2010) and with other body measurements such as antennal length and thorax width (Arango 1996). Furthermore, wing length is a better indicator of body size than wet or dry mass because it does not vary with lipid or water content. Finally, wing length is easier, faster, and less expensive to obtain than lean body mass, the other commonly used measure of lepidopteran body size (Miller 1977), and it does not require killing the butterflies or prolonged handling or storage times.

Although protocols for forewing measurement have been described three times (Beall & Williams 1945; Donham & Taylor 1996; Oberhauser et al. 2009), the authors used different landmarks, provided few specific details of how measurements were to be made, and did not provide an alternative measure that could be used when wing tips are missing or frayed. Furthermore, none of the existing protocols have been consistently adopted, thereby limiting the value of comparative wing length data.

Our review of the literature shows considerable variation in the device, precision of measurement, landmarks, and methods used to measure forewing length (Table 1). Rulers, calipers, an optical device [see Williams 1943], and computer programs that measure scanned images have been used. The landmarks employed include: from wing tip to wing tip (Dively *et al*. 2004), hind wing length (Herman 1988; Herman *et al*. 1989), wing area (Altizer & Davis 2010; Davis *et al*. 2007; Davis 2009), and the longest straight-line distance from the forewing base to the apical margin (forewing length; Table 1), which has been the most common measure. We found six distinct base points from which forewing straight-line measurements originated, and often no landmarks were recorded (Table 1). We also found differences in how this measurement was taken, including from left or right forewings (sometimes with the left and right forewings averaged), from dorsal or ventral surfaces, from intact butterflies, and from wings that had been removed from the body. Measurements have been taken while the butterflies were hand-held or lying on a surface, from live or dead individuals, and from dried specimens. Finally, we found no records of how damaged or worn wings were measured.

To reduce inconsistencies both within and among studies, we describe a specific forewing measurement method based on well-defined morphological landmarks that can easily be learned and used by both scientists and amateurs. Additionally, we address five questions regarding the protocol: 1) Should males and females be analyzed separately?; 2) Is recorded wing length affected by who makes the measurement?; 3) Does wing length decrease due to water loss from drying and, if so, does this bias measurements taken using the standard protocol?; 4) Do ruler measurements to the nearest 1 mm differ from caliper measurements to the nearest 0.1 mm?; and 5) Can forewing cell length be used to estimate total forewing length when neither forewing can be measured due to wing tip fraying or damage? Our standardized protocol describes both the method of measurement and recommendations based on our answers to these five questions. Widespread adoption of this protocol would greatly increase the comparative value of monarch wing length measurements by increasing repeatability of measurements within and among observers (Francis & Mattlin 1986; Bailey & Byrnes 1990; Gordon & Bradtmiller 1992; Ulijaszek & Kerr 1999; Harris & Smith 2009).

MEASUREMENT PROCEDURE

The right forewing should be measured (Fig. 1) unless it is deformed or part of the wing tip is frayed or missing, in which case the left forewing should be measured. If neither forewing can be measured directly, the right forewing cell length should be measured (see below). The forewing cell length can be used in the regression equation (provided in the Results) to estimate total wing length of the damaged or deformed forewing. The method used to measure the right forewing cell length is described below.

Regardless of handedness, one should hold the butterfly in the right hand with the thorax sandwiched between the thumb and forefinger and the right wings facing upward, as illustrated in Fig. 2. A firm but gentle pressure on either side of the thorax provides a stable platform for ruler placement while forcing the wings into the closed position, greatly reducing the likelihood of escape when measuring live butterflies.

Figs. 3 and 4 show the precise landmarks and proper line of forewing measurement. To measure the right forewing, locate the single white spot on the forewing at the forewing-thorax junction. This white spot, magnified and labeled as white spot #1 in Fig. 4, can be easily differentiated from the several white spots on the thorax by gently grasping the forewing along the leading margin and rotating the wing slightly upward toward the head of the butterfly. The correct spot moves with the wing. Be sure NOT to include any part of the thorax in the measurement. Place a transparent ruler so that the face of the ruler lies against the wing surface (the measurement is read through the backside of the ruler). The 1 cm rule line should be carefully aligned over the side of the white spot that is closest to the thorax. This ensures that the entire spot is included in the measurement (i.e., do NOT measure from the center of the spot). We suggest carefully placing a piece of masking tape at the 1 cm mark of the ruler to make it easier to align this mark with the base landmark on the wing. Once the basal point is set, gently rotate the ruler's leading edge across the margin of the apex (wing tip) until the maximum length is located. Pivoting the ruler on the basal landmark while gently pressing it against the wing surface requires considerable dexterity with the left hand. Be sure not to press so hard that the wing surface bows. This pivoting technique should be practiced until the observer obtains the same values in repeated measurements.

A clear ruler must be used for two reasons. First, it is important to stabilize the ruler by laying it across the forewing surface. This helps maintain the 1 cm mark at the precise base position while rotating the leading edge of the ruler along the wing tip and helps keep live butterflies immobilized. The ruler should not be held above the front margin of the wing because it cannot be braced from this position and butterflies can more easily escape. Secondly, when the right forewing is facing up, the wing tip from which the measurement is taken faces to the left (see Fig. 1). Thus, the measurement must be read from the right to the left. Because rulers read from left to right, the only way to achieve both proper ruler stabilization and the ability to take the measurement from right to left, is to read the measurement through the backside of a clear ruler. Therefore, the front surface of the ruler should be placed against the right forewing surface so that the numbers face right side up but backwards (see Fig. 2).

The total length of the red plus green arrows in Fig. 3 indicates the correct line of measurement. Once the leading edge of the ruler is correctly positioned on the wing tip, one must check to be sure that the 1 cm mark did not shift away from the basal landmark while rotating the ruler. Readjust if necessary. Using 1 cm rather than 0 cm as the starting point for measurement increases repeatability of measurements, but one must remember to subtract 1 cm before recording the length. Record all measurements to the nearest whole mm. If a measurement appears to fall exactly between two consecutive millimeter lines (i.e., exactly at 0.5 mm), then, following our standard protocol, it should be rounded to the nearest EVEN whole number. This method produces unbiased rounding.

When both forewing tips are damaged, the forewing cell may be measured instead (this should be noted). The forewing cell is enclosed by a series of wing veins and is forked, as outlined in yellow on Fig. 3. The cell length is measured from the same base landmark used to measure the total forewing length (from the base of the green arrow in Figs. 3 and 4). The distal landmark is defined as the intersection of the two wing veins that create the tip of the distal fork that sits farthest from the apical (front) margin of the wing (noted by the left tip of the green arrow in Fig. 3). Black wing scales surrounding the wing veins make this point difficult to locate precisely, thus requiring both good lighting and practice. When possible, we recommend observing the wing under a large self-standing magnifier for this purpose. The green arrow in Fig. 3 marks the proper line of measurement. Cell measurements should be taken to the nearest 0.5 mm in order to use our regression equation (see Results) to estimate total wing length to the nearest 1 mm.

METHODS

Collecting and handling the butterflies

A total of 56 wild adult monarchs were netted at Newport State Park, Door Co, WI, on 16 Jun 2009. They were immediately placed individually in glassine envelopes and stored with ice packs, and the butterflies were killed within three hours by placing them in a standard cooler containing dry ice. We refer to 'fresh' butterflies as either alive or dead but not yet dried for storage or chemical analyses. In this study, all fresh butterflies were measured dead but un-dried. They were shipped on ice among the observers and then stored inside a storage bag in a freezer. For dry measurements, butterflies were dried in a forced draft oven at 60˚C for 16 hours, the typical drying regime used for lipid analysis (Brower *et al*. 2006). Dried butterflies were shipped and stored in their original glassine envelopes inside a large plastic storage bag with desiccant.

Whether measured fresh or dry, butterflies were removed from their respective storage containers in batches of five, measured, and returned immediately. Following the measurement protocol described in the section above, all measurements were taken with wings intact on the body. With the exception of Question 2, which addresses differences in measurement among observers, a single experienced observer (TVH) took all measurements to avoid introducing inter-observer differences. All measurements

Fig. 1. Monarch in the position from which wing length measurements are taken: wings together, right wing facing up, and butterfly facing to the right.

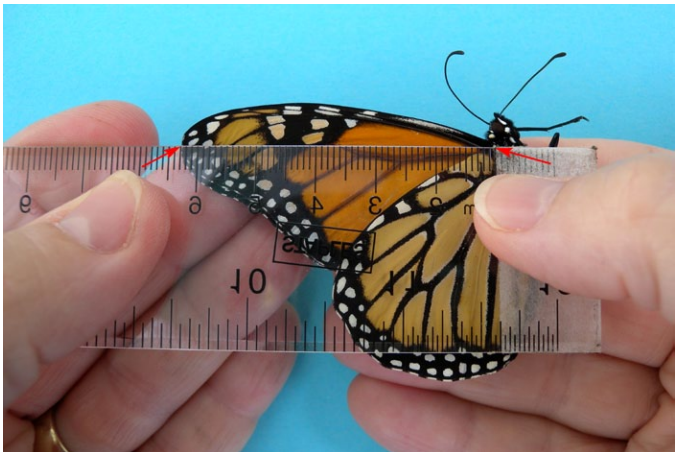


Fig. 2. Proper hand and ruler positioning for taking standardized right forewing length measurements. The thorax is sandwiched between the right forefingers and the ruler, which is braced by the right thumb. The left hand is used to support the wing tip while rotating the ruler to determine the maximum length. The face of the ruler is placed against the wing surface so that the measurement must be taken through the backside of the ruler (numbers are facing backwards). All standardized wing length measurements are taken from this position. The wing length is measured as the straight-line distance between the two red arrows. In this example, the total forewing length is 53 mm (63 mm minus 1 cm because the measurement is always taken from 1 cm ruler mark rather than 0). Proper hand and ruler positioning and measurement protocol are described in the text.

were taken independently, i.e., without knowledge of previous measurements. Data were analyzed using PASW Statistics 18 (PASW 2010).

1. Should males and females be analyzed separately?, and
2. Is wing length affected by who makes the measurements?

Questions one and two were answered using the same data set. For these two questions only, the left forewing was measured rather than the right forewing as specified in the standard protocol. Two observers measured 49 fresh butterflies, 23 males and 26 females, at 1 mm precision three times each, for a total of 294 measurements. To test for the effect of sex and observer on wing length, we ran a model I ANOVA with sex and observer as fixed factors on the average of the three repeated measurements for each butterfly. Averaging the replicate values reduces measurement error for assessing the significance of the main factors (Yezerinac *et al*. 1992).

Technical error of measurement (T.E.M.), an estimate of absolute measurement error expressed in the units of measurement (Mueller & Martorell 1988), was used to quantify and compare intra-observer and inter-observer variation in measurement. Intra-observer T.E.M. was calculated as the square root of the variation of the first two repeated measurements of individual butterflies averaged across all butterflies (Mueller & Martorell 1988 and references within). Only the first two measurements were used so that intra-observer error could be compared directly to inter-observer error. Inter-observer T.E.M. was calculated using the first set of measurements taken by each of our two observers (see WHO Multicentre Growth Reference Study Group 2006). A variance ratio test was used to compare the variances in the measurements of the two observers (Zar 1996).

Because the impact of measurement error depends on how much of the overall phenotypic variation in wing length is introduced by the process of measurement, we also report percent error (percentage of total phenotypic variation in wing length due to T.E.M.). We also report the reliability coefficient, R (R= 1 – (variance due to T.E.M.)). It represents the proportion of total variance that is due to true variation in forewing length.

3. Does wing length decrease due to water loss from drying?

To better detect changes in wing length due to drying, we altered the standard measurement protocol; ruler measurements were taken to the nearest 0.5 mm instead of the nearest 1.0 mm. The right forewing of 31 females and 24 males were measured three times fresh and three times after drying the butterflies, for a total of 330 measurements. We analyzed the averages of the replicate measures using a model 1 ANOVA with water status (fresh vs. dry) and sex as fixed factors. In addition, to assess whether there was an influence of butterfly size on the amount of shrinkage, we regressed the difference between fresh and dry wing length against fresh length.

4. Do ruler measurements to the nearest 1 mm differ from caliper measurements to the nearest 0.1 mm?

Because we found no effect of sex on measurement, we measured only females for this analysis. Two repeated measures of the right forewing length were made from 31 dried females using a ruler to the nearest 1.0 mm and electronic digital calipers (Mitutoyo Digimatic 150 mm/6 in) to the nearest 0.1 mm, for a total of 124 measurements. We compared the effect of instrument on wing length measurement using a paired t-test on the averages of the two wing measurements. T.E.M. was calculated from the two repeated measures to compare intra-observer variability in re-measures when measurements were taken with a ruler compared to calipers.

5. Can forewing cell length be used to estimate total forewing length when neither forewing can be measured due to wing tip fraying or damage?

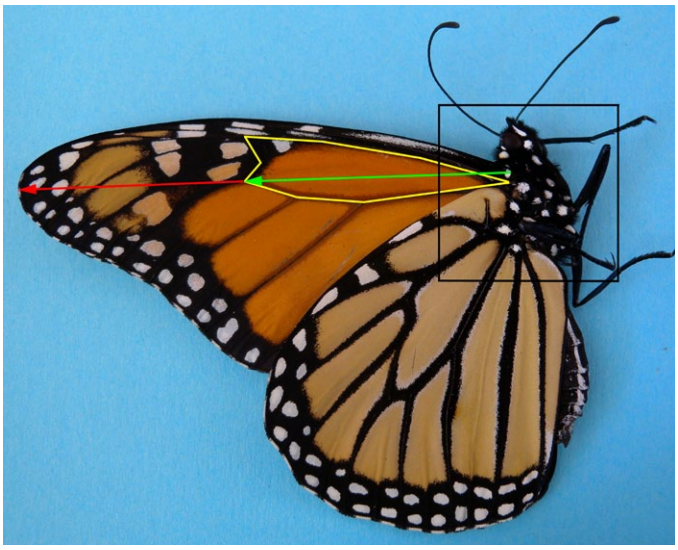


Fig. 3. Correct lines of wing measurements: Total forewing length is measured as the longest straight-line distance from the wing base to the wing tip (green + red arrows). The proper line of measurement depends on the wing tip shape but approximately bisects the forewing cell (outlined in yellow). Forewing cell length (green arrow only) is measured from the wing base to the tip of the lower prong of the forewing cell, delineated by the wing veins (noted by yellow outline). Notice that both measurements start at the far right (as pictured) edge of the white spot at the base of the forewing. See Fig. 4 for an enlarged view of the base landmarks used for both measurements.

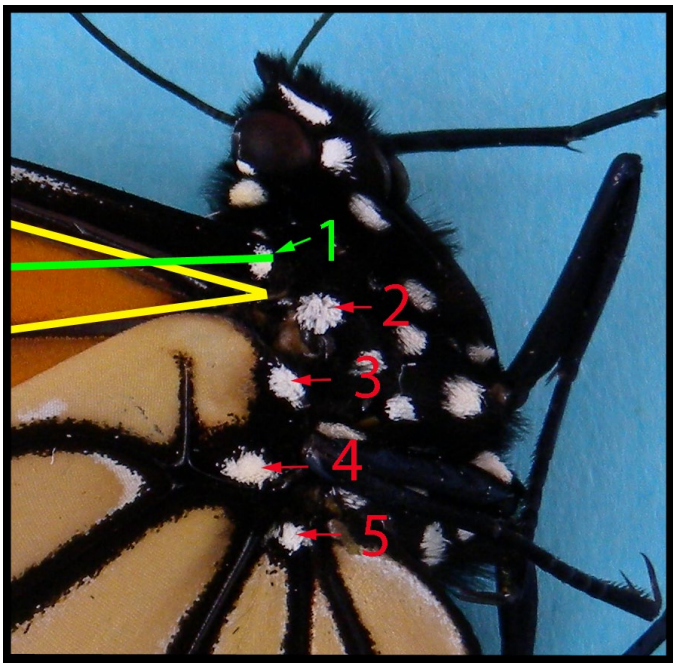


Fig. 4. Magnification of Fig. 3. Both monarch forewing length and forewing cell length are measured from the same base landmark (white spot #1). It is important to first differentiate the forewing white spot from the nearby white spot on the thorax (spot # 2) and the 3 white spots on base of the hind wing (spot #s 3, 4, and 5). To locate the correct spot, rotate the forewing slightly toward the head while holding the butterfly at the base of the hind wing. Only the spot on the base of the forewing will move. It is important to include the entire width of the landmark spot in the measurement as indicated. The right end of the green line of measurement (noted by green arrow) marks the exact location from which both the forewing length and forewing cell lengths originate. (The yellow lines delineate the forewing cell as shown in Fig. 3 and are not used as a landmark; see text for further explanation).

Ruler measurements were taken to the nearest 0.5 mm instead of to the nearest 1.0 mm for this question. Two repeated measures were taken from both the right forewing and forewing cell of 31 dried females, for a total of 124 measurements. Using the mean of the two measurements from each butterfly, total right forewing length and right forewing cell length were analyzed by linear regression.

RESULTS

1. Should males and females be analyzed separately?

Male forewing length was on average significantly greater than female wing length (mean = 52.22 mm, s.d. = 1.40 mm for males; mean = 50.95 mm, s.d. = 1.82 mm for females; ANOVA, $F_{1,94} = 15.090$, $p < 0.001$) (Table 2). Therefore, wing length should be analyzed separately for males and females.

2. Is wing length affected by who makes the measurements?

Based on the averages of the three repeated measurements, the two observers measured differently ($F_{1,94} = 4.878$, $p = 0.030$; see Table 2), while there was no significant interaction between observer and the sex of the butterfly ($F_{1,94} < 0.001$, $p = 0.992$). The two observers also differed in the variance of their repeated measurements ($F_{48,48} = 1.823$, $p < 0.05$), a difference reflected in the calculation of their T.E.M. values (observer 1 = 0.23 mm; observer 2 = 0.40 mm; see Table 2). Intra-observer measurement error ranged between 2 and 6% of total population variance for our two observers. Thus, between 94 and 98% of the total variance is true variation in wing length (coefficient of reliability, R = 0.94-0.98).

Inter-observer T.E.M. was 0.61 mm, so variance due to differences between our two observers was seven times larger than intra-observer 1 variance and two times larger than intra-observer 2 variance. In summary, observers differed in the variability of their repeated measurements and recorded significantly different wing lengths for the same butterflies. Butterfly sex did not affect observer differences in measurement. Finally, measurements taken by two observers varied much more than those taken by a single observer.

3. Does wing length decrease due to water loss from drying?

The mean wing length decreased from 51.78 in the fresh state to 51.42 mm when dried ($F_{1,106} = 11.249$, $p = 0.001$), representing an average decrease of 0.36 ± 0.05 mm due to shrinkage upon drying. This shrinkage is very small, representing $0.7\% \pm 0.1\%$ (mean and 95% C.I.) of the fresh wing length. The amount of shrinkage did not differ between the sexes ($F_{1,106} = 0.031$, $p = 0.860$), nor did shrinkage differ with wing length (fresh-dry difference regressed on wing length, $F_{1,53} = 0.001$, n.s.). Although the amount of shrinkage due to drying is less than the precision level (0.5 mm) used to measure the butterflies, combining fresh and dry wing lengths would increase variance and decrease the sensitivity of the analysis.

4. Do ruler measurements to the nearest 1 mm differ from caliper measurements to the nearest 0.1 mm?

Table 1. A non-exhaustive literature survey of measurements of monarch wing size. A dash (-) means that the information was not reported. Device: R = ruler, D = digital, C = calipers, O = optical device. Side: R = right, L = left, AV = average of left and right. Precision: Precision level of measurement (mm). Landmarks: a = dorsal surface from the proximal costal forewing corner to the most distant point in the wing apex, a* = some measurements were taken using a special optical device without disturbing museum specimens, b = ventral forewing length measured from the wing attachment to the apex of wing, c = white spot on forewing base to the apex of wing, d = white spot on thorax to the apex of wing, e = thorax to longest extension, f = distal tip of left forewing to distal tip of right forewing, HW = hind wing measured, Area = area measured. Comparison: 1 = sexes, 2 = correlated physical traits, 3 = correlated behaviors, 4 = larval rearing conditions, 5 = differences across overwintering season, 6 = population comparisons, 7 = generation comparisons, 8 = differences across years at single location, 9 = differences across fall migration phase at single location, 10 = other comparison. We assume a single observer made the measurements when a paper had a single author. The table shows extreme heterogeneity in measurement protocols published between 1945 and 2010.

Source	Device	Side	Precision	Land-marks	Multiple Observers	Comparison
Alonso et al. 1997	R	R	0.5	b	No	2,3,5
Altizer 2001	C	-	-	b	No	10
Alitzer & Oberhauser 1999	C	-	-	b	-	2
Altizer et al. 1999	C	-	-	b	-	10
Alonso et al. 1997	R	R	0.5	b	No	2,3,5
Altizer & Davis 2010	D	R	-	b	-	1,2,6,5,10
Altizer et al. 2004	-	-	-	b	-	10
Arango 1996	C		0.05	c	No/Yes	2,4,6,7,8
Beall 1946	-	R	1	a	No	1,2,6,7,8,9,5,10
Beall & Williams 1945	O + R	R	1	a*	Yes	6,9
Borland et al. 2004	-	-	-	b	Yes	1,2,8,9,10
Bradley & Altizer 2005	D	-	-	Area	-	10
Brindza et al. 2008	C	L	0.1	b	No	1,6,8,9
Brower et al. 2006	R	-	0.5	-	-	8
Brown & Chippendale 1974	-	-	-	-	-	2
Calvert & Lawton 1993	-	-	-	-	-	1,2,5,9
Davis 2009	D	-	-	Area	No	2
Davis et al. 2007	D	AV	.	Area	-	2,3
Dively et al. 2004	-	.	-	f	-	4
Dockx 2002, 2007	-	R	-	c	No	1,4,6
Eanes 1978	-	-	-	-	No	2
Frey et al. 1998	-	AV	1	b	-	1,2,3,10
Gibo & McCurdy 1993	-	R	0.5	-	-	9
Herman 1988	R	-	0.5	HW	No	1,7
Herman et al. 1989	R	-	0.5	HW	-	1,5,6,7
James 1984	-	R	-	-	No	1,7,8
Jesse & Obrycki 2000, 2004	-	R	-	d	-	4
Knight 1998	R	R	0.5	c	No	1,6
Lavoie & Oberhasuer 2004	C	-	0.1	b	-	4
Leong et al. 1993	R	R	1	-	-	1,2,3
Leong et al. 1995	R	-	1	-	-	3,5
Levine et al. 2003	-	-	1	e	-	5
Lindey & Altizer 2009	D	-	.	Area	-	4
Malcolm et al. 1989	-	R	-	-	-	1,4
Oberhauser 2004	-	-	0.1	-	No	2
Oberhauser & Frey 1999	-	-	-	-	-	3
Solensky & Oberhauser 2004	-	-	-	-	-	3
Solensky & Oberhauser 2009	-	-	-	Area	-	2,3
Tuskes & Brower 1978	-	R	-	-	-	1,5
Van Hook 1993, 1996	-	-	0.5	b	No	1,2,3,4,5
York & Oberhauser 2002	-	-	-	b	-	4

Measurements taken with a ruler to the nearest 1 mm and with calipers to the nearest 0.1 mm gave similar descriptive results; the overall mean right forewing lengths, calculated from the means of two repeated measures for all 31 butterflies, were 50.98 mm at the 0.1 mm level of precision and 51.06 mm at the 1.0 mm level (Table 3). Based on a paired comparison of ruler and caliper averages for each butterfly, there was no significant effect of measurement device on forewing length measurement ($t_{31} = 1.669$, $p = 0.106$). The T.E.M. for ruler measurements was 0.25 mm compared to 0.17 mm caliper measurements (see Table 3). This difference in variability of repeated measurements was not statistically significant (variance ratio test: $F_{30,30} = 1.0220$, $p > 0.05$; Table 3), and measurement error using either device represented less than 2% of the total variance. Thus, we conclude that measurement of wing length by ruler with 1.0 mm precision is sufficient for most studies (but see discussion).

Post-hoc statistical power analysis (Soper 2011) of our forewing length data measured with a ruler to the nearest 1 mm showed that sample sizes of 30 individuals (15 in each group for 2 factor analyses) provide sufficient power to detect large and medium effects (power = 0.86 to 0.92 and 0.48 to 0.61 for one-tail and two-tail hypotheses for large and medium effects, respectively, at alpha = 0.05). Much larger sample sizes would be necessary to detect very small forewing differences among populations.

5. Can forewing cell length be used to estimate total forewing length when neither forewing can be measured due to wing tip fraying or damage?

Using the mean of the two repeated measures for the right forewing length and right forewing cell length, we found these measurements to be correlated significantly ($R^2 = 0.75$, $p < 0.01$, $n = 62$; Fig. 5). We conclude that the forewing cell length, based on the mean of two repeated measures to the nearest 0.5 mm, can be used as an adequate alternative measure to estimate forewing length when neither right nor left forewing can be measured directly. The regression equation was: forewing length in mm = 1.508 (forewing cell length mm) + 10.744 mm.

DISCUSSION

Sex

We found that male wing length was on average 1.27 mm greater than female wing length. Our result was consistent with many (e.g. Beall 1946; Herman 1988; Calvert & Lawton 1993; Van Hook 1993, 1996; Knight 1998; Oberhauser & Frey 1999; Borland *et al*. 2004; Dockx 2002, 2007; Brindza *et al*. 2008; Altizer & Davis 2010) but not all (e.g. Tuskes & Brower 1978; James 1984; Frey *et al*. 1998; Leong *et al*. 1993; Knight 1998; Malcolm *et al*. 1989; Dockx 2002, 2007) findings based on various monarch populations. Although our literature review showed that the sexes are usually considered separately (Table 1), this was not always the case (e.g. James 1984 and some comparisons in Brindza *et al*. 2008). The standard protocol therefore includes the provision that the sex always be recorded and that males and females be considered independently in analyses of wing length.

Table 2. Summary statistics for forewing length measurements of our Door County, WI sample. Sex and observer effects are shown. All measurements were taken with a ruler to the nearest 1 mm. Mean and SD are based on 2 repeated measurements for each butterfly. Technical error of measurement (TEM) and percent measurement error (%ME) are based on the variance of 2 repeated measurements for each observer (intra-observer) and the variance of the 1st measurement for 2 observers (inter-observer).

	Intra-Observer								Inter-Observer	
	Observer 1				Observer 2				Observer 1 vs 2	
	Mean	SD	TEM	% ME	Mean	SD	TEM	% ME	TEM	% ME
Males	52.58	1.36	0.21	2.3	51.86	1.38	0.39	7.7	0.66	21.0
Females	51.31	1.80	0.24	1.8	50.59	1.80	0.42	5.6	0.57	10.4
Both	51.91	1.71	0.23	1.7	51.18	1.75	0.40	5.5	0.61	12.8

Mean comparison of the sexes, $p < 0.001$
Mean comparison of the observers, $p = 0.030$
Variance differs between observers, $p < 0.05$

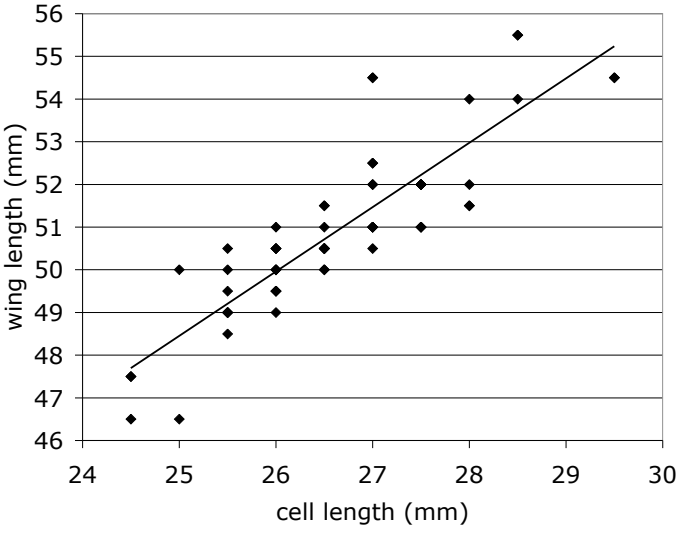
Table 3. The effect of measurement device on the forewing descriptive statistics, technical error of measurement (TEM), and percent of total variance due to repeated measures (% ME) for 31 dried females measured by a single observer with calipers at 0.1 mm and a ruler at 1.0 mm. Mean and SD are based on the averages of the two repeated measurements for each butterfly, while TEM and % ME are based on the variance of the two repeated measurements.

Measurement Procedure	Mean [a]	SD	TEM	% ME [b]
Calipers at 0.1 mm	50.98	1.91	0.17	0.8
Ruler at 1.0 mm	51.07	1.89	0.25	1.8

[a] Comparison of means, $p = 0.106$, n.s.

[b] Variance ratio test of variance between measurement devices: $p > 0.50$, n.s.

Fig. 5. Regression of right forewing length vs. right forewing cell length: $y = 1.508x + 10.744$ ($r^2 = 0.747$). A single observer measured total forewing length and cell length two times each to the nearest 0.5 mm using a ruler, and all measurements were independent. The averages of two repeated measurements were used for the regression.



Intra- and Inter-observer error using the Standard Measurement Protocol

Measurement error is inversely related to quality of the data, and standardization of the measurement procedure is the most effective way to minimize such error (Ulijaszek & Kerr 1999). Using the standard protocol, our two observers both showed 'excellent' agreement in their repeated measurements (measurement errors less than 10%: see Stokes 1985; Perini *et al*. 2005; WHO Multicentre Growth Reference Study Group 2006). However, they differed in their measurement errors: 2% and 6% based on the combined sex sample. Consistent with the literature, the observer who was most experienced using the standard protocol showed a higher level of repeatability (Gordon & Bradtmiller 1992; Yezerinac *et al*. 1992; Tong *et al*. 1998; Ulijaszek & Kerr 1999; Kania 2004; references in Perini *et al*. 2005). We conclude that with a single observer, single measurements with a ruler to the nearest 1 mm are adequate, but practice using the standard protocol before data collection begins is necessary to minimize error.

The measurements of our observers were consistently different. This difference, combined with differences in the size of their measurement errors, increased inter-observer error compared to intra-observer error. Larger variation in measurements among observers compared to repeated measurements by a single observer is a common finding even when the measurement protocol is standardized (Yuan *et al*. 2004; Geeta *et al.* 2009; Harris & Smith 2009; Muñoz-Muñoz & Perpiñán 2010; but see Palmeirim 1998; Ulijaszek & Kerr 1999).

The effect of any increased variation caused by using multiple observers depends on the total variance in wing length of the population measured. Our data illustrate this: T.E.M. was similar for females and males. However, these errors varied over two-fold in total length variation (from 10 to 21% for inter-observer error, respectively, Table 2). Even though measurement error would be larger for males than females, the coefficients of reliability, 0.90 and 0.79 for females and males respectively, still indicate from 'excellent' to 'substantial' agreement between our measurers (WHO Multicentre Growth Reference Study Group 2006). However, because percent measurement error was consistently and substantially smaller for single observers compared to multiple observers (Table 2), we recommend that a single measurer be used whenever possible to minimize relative bias and variation in measurements (Measey *et al*. 2003; Perini *et al*. 2005).

Our estimates of relative bias and inter-observer variation in measurement may have been inflated because our observers did not calibrate their measurements on the same butterflies prior to the study. The literature suggests that both variation in measurement and relative bias could be reduced perhaps well beyond those found in this study if observers calibrate their measurements after adequate training using the standard protocol (see Gordon & Bradtmiller 1992; Kouchi *et al*. 1999; WHO Multicentre Growth Reference Study Group 2006). To calibrate measurements among observers, all measurers in a single study or at a single monitoring site and date should work as a group to compare their measurement techniques and resulting measurements on a single sub-set of butterflies to identify subtle differences and then work together until everyone is satisfied that they have minimized those differences. When observers change through time, as is common in monitoring studies, established observers should help to train the next group.

Although inter-observer error cannot be factored out during statistical analyses when single measurements are used (Harris & Smith 2009), we recommend against taking the mean of repeated measurements or dramatically increasing sample sizes because they are not necessary. This is because repeated measurements do not reduce measurement differences among observers, and increased sample sizes can magnify them (Palmeirim 1998). Instead we recommend putting the time it would take to obtain duplicate measures on all butterflies into proper training, practice, and calibration before collecting data. The effects of any remaining differences in measurement can be mitigated when multiple observers are used within a single study by dividing each of the butterfly groups of interest among all of the observers. For example, if the question is whether wing length differs between coastal and inland migratory monarchs, each observer should measure both coastal and inland.

We encourage monitoring groups to teach the standard protocol and calibration methods. However, calibration among sites and across time will often be unfeasible. We therefore make the following additional recommendations. First, the names and handedness of observers should be reported (see Helm & Albrecht 2000), and raw data should be archived so they can be accessed for statistical comparisons. Second, after training and calibrating, a sub-sample of at least 20 butterflies should be measured independently twice by all observers. Intra-observer and inter-observer T.E.M. should be reported, and reassessments should be made at regular intervals (Yezerinac *et al*. 1992; see Mueller & Martorell 1988 for methodology to determine T.E.M.). Third, if inter-observer error is greater than 10% of total variation in wing length attempts should be made to further reduce differences in measurement technique among observers. When inter-observer error is 10% or less or after attempts have been made to reduce observer differences, the first of the paired measurements used to calculate observer error can be included in the overall data set to minimize resources and time needed to document reliability.

There is no way to assess true bias in measurement, and it would be impractical to compare relative bias across sites because the same butterflies cannot be used. However, relative bias can be reduced through calibration (Kouchi *et al*. 1999) and is included in inter-observer error. Therefore, when inter-observer error remains high or unknown after training, calibration, and practice, small significant differences based on samples measured by different observers should be viewed with caution (Palmeirim 1998).

Fresh vs. dry butterfly measurements

Measured wing length of fresh monarchs was on average 0.36 mm greater than measurements after they were dried, with no significant difference by sex in the amount of shrinkage. Shrinkage was small compared to the 1 mm precision level used in the standard protocol and represented less than 1% of the total fresh wing length. However, because all sources of

variation among measurements are cumulative, we recommend that observers measure wing length before drying when possible and note this. If fresh wing lengths must be compared to dried wing lengths, we suggest first estimating wet wing length from dry wing length by adding 0.7% to all dry wing lengths and interpreting small significant differences in wing lengths cautiously.

Water evaporation during long term freezing can presumably decrease wing length slightly in the same way as drying them for chemical analyses. Since the degree of drying and thereby shrinkage will vary, we recommend first drying such specimens for 16 hrs at 60 degree C, taking 'dry' wing length measurements, and then adding 0.7% to convert wing length measurements to fresh equivalents. We know of no empirical evidence or theoretical reason that wing length should differ between live and just frozen specimens. However, measurements on live specimens may be more variable (see below) because they are more difficult to measure.

When museum specimens are measured, not only have the wing lengths shrunk due to water loss, but the standard protocol must also be altered. Such measurements would not be easy to standardize across all observers because the dorsal surface has no white spots to serve as base landmarks; it is harder to differentiate thorax from wing on spread specimens; and specimens are so fragile that they usually cannot be touched. When wing lengths of museum specimens must be compared to those taken using the standard measurement protocol, we recommend using the following procedure. First, measure a set of fresh monarchs using the standard protocol. Then, after pinning and drying the butterflies at 60 degrees C for 16 hours, measure them again to create a regression of fresh vs. pinned (dried and spread) wing lengths. Finally, using the same measurement protocol and the same observer, measure the museum sample of interest and convert the measurements to fresh equivalents using the established regression.

Measurement device and sample size

We based the standard protocol on ruler measurements because the ruler is the most common measurement device reported in the literature, and its widespread use maximizes comparability of wing length measurements across long-term monitoring programs, citizen scientists, and researchers. Rulers are also easier to use and more affordable than calipers. Ruler measurements taken to the nearest 1 mm did not differ significantly from those taken with calipers to the nearest 0.1 mm, nor was measurement error significantly increased. Furthermore, error associated with ruler measurements (less than 2%) was considerably smaller than error introduced by different observers (10-21%).

A researcher's choice of sample size depends on the question at hand, population variation, the effect size of interest, confidence level needed, and measurement error. However, measurement error is included in the overall population variance when single measurements are taken, causing a slight loss in statistical power (Yezerinac *et al*. 1992). This can be countered by increasing sample size. Based on post-hoc statistical power analyses, we found that using sample size of at least 30 (15 in each sample when two population are compared) was sufficient to detect large and medium effects with single observers. We recommend using power analysis to assure adequate sample sizes, especially when small differences in wing length are important.

When different populations are measured by different observers, the results should be viewed with caution because any significant differences may be a product of the observers rather than the populations (Palmeirim 1998). When fewer than 30 butterflies are measured, two replicate ruler measures should be averaged to decrease measurement error. Furthermore, we recommend that calipers be used in studies whose goal is to detect very small differences (e.g., when assessing right-left wing asymmetry). Our results indicate that measurements of forewing length taken with calipers could be rounded to the nearest 1 mm for comparison to ruler measurements.

Computer-based measurements allow higher levels of measurement precision, may provide better intra-observer repeatability (see Muñoz-Muñoz & Perpiñán 2010), and may be preferred under some circumstances (e.g. Davis *et al*. 2007; Altizer & Davis 2010). However, this method is not suited for general use because it increases handling time, (thereby adding stress to live animals), requires special computer programs, and increases inter-observer biases because of calibration difficulties (see Muñoz-Muñoz & Perpiñán 2010). Because software differences make standardization of computer analyses across observers difficult, we suggest that ruler measurements be taken in addition to image-based measurements when measurements are to be compared to those taken by the standard protocol.

Side of body measured and total forewing wing length estimation

When left and right wing lengths are combined, small errors can result from true asymmetry in right and left wing lengths (Palmer & Strobeck 1986) or from biases resulting from how the two sides are measured (Arango 1996; Helm & Albrecht 2000). Therefore, following Beall & Williams (1945) and the prevailing trend in the existing monarch literature (Table 1), the standard protocol restricts measurement to the right forewing. However, the difference in right-left wing length is small compared to the 1 mm precision level used in the standard protocol (Arango 1996). It is important to measure all butterflies in a particular sample, even when the right forewing cannot be measured, because wing length may be correlated with other factors of interest, such as behavior, age, population source, etc. (Van Hook 1993; Oberhauser & Frey 1999; Oberhauser *et al*. 2009). The standard protocol therefore dictates that the left forewing be measured when the right forewing is not intact, and this should be noted.

When neither forewing can be measured, the right forewing cell length should be measured twice to the nearest 0.5 mm. The mean can then be used to estimate total forewing length based on our regression equation. Estimates from this regression are not as accurate as direct measurements, of course, so it depends on the questions being asked whether regressed estimates should be used in a study. (Measurers should note when estimates are used on their data sheets, and the percentage of estimated wing lengths should be reported.) Because the forewing cell length is more challenging to measure, excellent lighting conditions are

important, and magnification may be needed.

Although total hind wing length may provide a better regression equation for total wing length compared to forewing cell length, we recommend against using the hind wing length for two reasons. First, when the forewings are frayed or damaged, the hind wings are also often similarly worn (Leong *et al*. 1993). Secondly, the hind wing measurement is more difficult to standardize than forewing cell length because the wing margin is feathery and scalloped (see Fig. 3) and requires using a different base landmark.

Measurements taken on live monarchs

Measurement error may be higher when measurements are taken on live compared to dead animals because of the difficulty of holding them (Blackwell *et al*. 2006; personal experience). Therefore, when monarchs are being preserved, forewing length should be measured after they are killed. Measurements should be taken as soon as possible after collection to avoid possible shrinkage in wing length associated with water loss during long-term freezer storage (see above). However, one of the merits of using wing length to answer biological questions, especially relevant when citizen scientists gather the data, is that it does not require killing the butterflies. When monarchs are measured alive, Donham & Taylor (1996) suggest placing them inside clear envelopes to measure them, but we do not know how this method might influence variability or relative bias of measurements. Instead, we recommend cooling the butterflies by placing them into glassine collection envelopes that should be kept inside a sealed plastic bag stored in a cooler or until they are removed for measurement. Cooling keeps the butterflies quiet and prevents them from damaging their wings and using energy reserves.

If each butterfly must be measured immediately and then released, we suggest practicing the proper technique of holding the ruler over the body for taking measurements while firmly holding the thorax between the thumb and forefinger. Together, these techniques stabilize the ruler for more precise measurements while keeping the butterfly quiescent. Any increase in measurement error due to measuring live specimens likely increases variation rather than bias in measurement and therefore should not limit statistical analyses of the data as long as sample sizes are ≥ 30 individuals.

Summary of the standard forewing measurement protocol

We describe a standard protocol for forewing length measurements based on well-defined landmarks and specific methods for handling the butterfly and measurement device. This protocol provided 'excellent' repeatability of measurements when a single observer was used and from 'excellent' to 'substantial' repeatability with multiple observers (see WHO Multicentre Growth Reference Study Group 2006). Single ruler measurements to 1 mm should generally provide sufficient statistical power as long as sample sizes are robust (≥ 30) and a single observer is used. However, after assessing T.E.M., power analysis should be used to estimate appropriate sample size. When sample size is severely restricted, calipers or the mean of two repeated measures can be used to increase the power of statistical tests.

We emphasize the importance of proper training and sufficient practice before collecting data. A single observer should be used when possible, but when multiple observers are necessary, inter-observer repeatability may be increased beyond levels found in our study if all measurers work together to calibrate their measurements. Because observer error varies and cannot be factored out when single measurements are used, the names, handedness, and intra-observer and inter-observer T.E.M. should be reported, while the raw data should be archived. Observer bias can be mitigated by sub-dividing samples among the observers so that observers and factors of interest do not co-vary. When this is not possible, small significant differences should be viewed with caution due to unavoidable measurement differences among observers (relative bias).

The sex should be recorded along with forewing length and males and females analyzed separately. All of the butterflies in a sample should be measured; use the left forewing when the right forewing apical tip is frayed or missing and use the means of right forewing cell lengths measured twice to the nearest 0.5 mm and our regression equation to estimate the total wing length when neither forewing can be measured. Measurements should be taken before drying dead specimens, and live butterflies should be cooled before measuring.

We hope that monarch researchers and monitoring groups will adopt this standardized measurement protocol. Its widespread use will increase the comparability and usefulness of monarch wing length data. The general methods we used for standardization can be applied to all lepidopteran species to increase repeatability in wing length measurement.

ACKNOWLEDGEMENTS

We are grateful to Michelle Hefty for permission to sample monarchs in Newport State Park, WI and Ray Sullivan for field assistance. We also thank Zachariah Dietz, who provided important statistical guidance, and Gwen Kirschke, who assisted with the photographs. Barbara Cockrell generously assisted us by obtaining references. Two anonymous referees provided helpful comments that improved the manuscript. This work was supported by the October Hill Foundation (LPB) and by the Christian A. Johnson Fund of Hamilton College (EHW).

REFERENCES CITED

Alonso-Mejía, A., E. Rendon-Salinas, E. Montesinos-Patiño & L. P. Brower
1997. Use of lipid reserves by monarch butterflies overwintering in Mexico: Implications for conservation. *Ecological Applications* 7(3):934-947.

Altizer , S. M.
2001. Migratory behaviour and host-parasite co-evolution in natural populations of monarch butterflies infected with a protozoan parasite. *Evolutionary Ecology* 3(3):611-632.

Altizer, S. M. & A. K. Davis
2010. Populations of monarch butterflies with different migratory behaviors show divergence in wing morphology. *Evolution* 64(4):1018-1028.

Altizer, S. M. & K. Oberhauser
1999. Effects of the protozoan parasite *Ophryocystis elektroscirrha* on the fitness of monarch butterflies (*Danaus plexippus*). *Journal of*

Invertebrate Pathology 74(1):76-88.

Altizer, S. M., K. Oberhauser & K. A. Geurts
2004. Transmission of the protozoan parasite *Ophryocystis elektroscirrha* in monarch butterfly populations: Implications for prevalence and population-level impact. pp. 203-218. *In* Oberhauser, K. S. & M. J. Solensky (Eds.), The Monarch Butterfly: Biology and Conservation. Cornell University Press, Ithaca, New York.

Arango, N.
1996. *Stabilizing selection in migratory butterflies: A comparative study of queen and monarch butterflies.* Unpublished Masters Thesis, University of Florida, Gainesville, Florida, i-viii + 1-70.

Bailey, R. C. & J. Byrnes
1990. A new, old method for assessing measurement error in both univariate and multivariate morphometric studies. *Systematic Zool*ogy 39(2):124-130.

Beall, G.
1946. Seasonal variation in sex proportion and wing length in the migrant butterfly, *Danaus plexippus* L. (Lep. Danaidae). *Transactions of the Royal Entomological Society of London* 97:337-353.

Beall, G. & C. B. Williams
1945. Geographical variation in the wing length of *Danaus plexippus* (Lep. Rhopalocera). *Proceedings of the Royal Entomological Society of London* 20:65-76.

Blackwell, G. L., S. M. Bassett & C. R. Dickman
2006. Measurement error associated with external measurements commonly used in small-mammal studies. *Journal of Mammalogy* 87(2):216-223.

Borland, J., C. Johnson, T. Crumpton, T. Thomas, S. Altizer & K. Oberhauser
2004. Characteristics of fall migratory monarch butterflies, *Danaus plexippus*, in Minnesota and Texas. pp. 97-104. *In* Oberhauser, K.S. & M.J. Solensky, (Eds.), *The Monarch Butterfly: Biology and Conservation*. Cornell University Press, Ithaca, New York.

Bradley, C. & S. Altizer
2005. Parasites hinder monarch butterfly flight ability: implications for disease spread in migratory hosts. *Ecological Letters* 8: 290-300.

Brindza, L., L. B. Brower, A. K. Davis & T. Van Hook
2008. Comparative success of monarch butterfly migration to overwintering sites in Mexico form inland and coastal sites in Virginia. *Journal of the Lepidopterists' Society* 62(4):189-200.

Brower, L. P., L. S. Fink & P. Walford
2006. Fueling the fall migration of the monarch butterfly. *Integrative and Comparative Biology* 46(6):1123-1142.

Brown, J. J. & G. M. Chippendale
1974. Migration of the monarch butterfly, *Danaus plexippus*: energy sources. *Journal of Insect Physiology* 20:1117-1130.

Calvert, W. H. & R. O. Lawton
1993. Comparative phenology of variation in size, weight, and water content of eastern North American monarch butterflies at five overwintering sites in Mexico. pp. 299-307. *In* Malcolm, S. B. & M. P. Zalucki (Eds.), *Biology and Conservation of the Monarch Butterfly*. Publications of the Natural History Museum of Los Angeles County, Los Angeles, California.

Davis, A. K.
2009. Wing color of monarch butterflies (*Danaus plexippus*) in Eastern North America across life stages: Migrants are "redder" than breeding and overwintering stages. *Psyche*:1-5.

Davis A. K., N. Cope, A. Smith & M. J. Solensky
2007. Wing color predicts future mating success in male monarch butterflies. *Annals of the Entomological Society of America* 100(2):339-344.

Dively, G. P., R. Rose, M. K. Sears, R. L. Hellmich, D. E. Stanley-Horn, D. D. Calvin, J. M. Russo & P. L. Anderson
2004. Effects on monarch butterfly larvae (Lepidoptera: Danaidae) after continuous exposure to Cryl-Ab-expressing corn during anthesis. *Environmental Entomology* 33(4):1116-1125.

Dockx, C.
2002. *Migration of the North American Monarchs Danaus plexippus to Cuba*. PhD Dissertation, University of Florida, Gainesville, Florida, x + 119.
2007. Directional and stabilizing selection on wing size and shape in migrant and resident monarch butterflies, *Danaus plexippus* (L.), in Cuba. *Biological Journal of the Linnean Society* 92:605-616.

Donham, P. & O. R. Taylor
1996. The Great Dragonfly Challenge: Measuring Monarchs. *Dragonfly Magazine Sept/Oct* (1):21-23.

Eanes, W. F.
1978. Morphological variance and enzyme heterozygosity in the monarch butterfly. *Nature* 276:263-264.

Francis, R. I. C.C. & R. H. Mattlin
1986. A possible pitfall in the morphometric application of discriminate analysis: measurement bias. *Marine Biology* 93:311-313.

Frey, D., K. L. H. Leong & E. Peffer, R. K. Smidt & K. Oberhauser
1998. Mate pairing patterns on monarch butterflies (*Danaus plexippus* L.) at a California overwintering site. *Journal of the Lepidopterists' Society* 52(1):84-97.

Geeta, A. H. Jamaiyah, N. M. Safiza, G. L. Khor, C. C. Kee, A. Z. Ahmad, S. Suzana, R. Rahmah & A. Faudzi
2009. Reliability, technical error of measurements and validity of instruments for nutritional status assessment of adults in Malaysia. *Singapore Medical Journal* 50(10):1013-1019.

Gibo, D. L. & D. A. McCurdy
1993. Lipid accumulation by migrating monarch butterflies (*Danaus plexippus* L.). *Canadian Journal of Zoology* 71:76-82.

Gordon, C. C. & B. Bradtmiller
1992. Interobserver error in a large scale anthropometric survey. *American Journal of Human Biology* 4:253-263.

Harris, E. F. & R. N. Smith
2009. Accounting for measurement error: A critical but often overlooked process. *Archives of Oral Biology* 54 *Suppl*. 1:107-117.

Helm, B. & H. Albrecht
2000. Human handedness causes directional asymmetry in avian wing length measurements. *Animal Behavior* 60:899-902.

Herman, W. S.
1988. Body weight and wing length changes in Minnesota populations of the monarch butterfly. *Journal of the Lepidopterists' Society* 42:32-36.

Herman, W. S., L. P. Brower & W. H. Calvert
1989. Reproductive tract development in monarch butterflies overwintering in California and Mexico. *Journal of the Lepidopterists' Society* 43(1):50-58.

James, D. G.
1984. Phenology of weight, moisture and energy reserves of Australian monarch butterflies, *Danaus plexippus*. *Ecological Entomology* 9:421-428.

Jesse, L. C. H. & J. J. Obrycki
2000. Field deposition of Bt transgenic corn pollen: lethal effects on the monarch butterfly. *Oecologia* 125:241-248.
2004. Survival of experimental cohorts of monarch larvae following exposure to transgenic corn pollen and anthers. pp. 69-75. *In* Oberhauser, K. S. & M. J. Solensky (Eds.), *Monarch Butterfly: Biology and Conservation*. Cornell University Press, Ithaca, New York.

Kania, W.
2004. Observer error in measurements of nestling wing length in small passerines. *The Ring* 26(2):79-87.

Knight, A.
1998. *A population study of monarch butterflies in North-Central and South Florida.* Unpublished Masters Thesis, University of Florida, Gainesville, Florida, viii + 125.

Kouchi, M., M. Mochimaru, K. Tsuzuki & T. Yokoi
1999. Interobserver errors in anthropometry. *Journal of Human Ergology* 28:15-24.

Lavoie, B. & K. S. Oberhauser
2004. Compensatory feeding in *Danaus plexippus* (Lepidoptera: Nymphalidae) in response to variation in host plant quality. *Environmental Entomology* 33(4):1062-1069.

Leong, K. L. H., D. Frey, D. Hamaoka & K. Honma
1993. Wing damage in overwintering populations of monarch butterfly at two California sites**.** *Annals of the Entomological Society of America* 86(6):728-733.

Leong, K. L. H., E. O'Brien, K. Lowerisen & M. Colleran
1995. Mating activity and status of overwintering monarch butterflies (Lepidoptera: Danainae) in Central California. *Annals of the*

Entomological Society of America 88:45-50.

Levine, E., O. Byron-Cooper, M. Desch-O'Donnell & C. Metzinger
2003. Interindividual variation in mitochondrial enzyme activity in male monarch butterflies, *Danaus plexippus* L. (Nymphalidae). *Journal of the Lepidopterists' Society* 57(4):284-290.

Lindsey, E. A. & S. Altizer
2009. Sex differences in immune defenses and response to parasitism in monarch butterflies. *Evolutionary Ecol*ogy 23:607-620.

Malcolm, S. B, B. J. Cockrell & L. P. Brower
1989. Cardenolide fingerprint of monarch butterflies reared on common milkweed, *Asclepias syriaca* L. *J. Chemical Ecology* 15(3):819-853.

Measey, G. J., J. Bernardo Silva & M. Di-Bernardo
2003. Testing for repeatability in measurements of length and mass in *Chthonerpeton indistinctum* (Amphibia: Gmnophiona), including a novel method of calculating total length of live Caecilians. *Herpetological Review* 34(1):35-39.

Miller, W. E.
1977. Wing measure as a size index in Lepidoptera: The family Olethreutidae. *Entomological Society of America* 70(2):253-256.
1991. Body size in North American Lepidoptera as related to geography. *Journal of the Lepidopterists' Society* 45:158-168.

Mueller, W. H. & R. Martorell
1988. Reliability and accuracy of measurement. pp. 83-86. *In* Lohman, T. G., Roche, A. F. & Martorell, R. (Eds.), *Anthropometric Standardization Reference Manual*, Human Kinetics Books, Champaign, Illinois.

Muñoz-Muñoz, F. & D. Perpiñán
2010. Measurement error in morphometric studies: comparison between manual and computerized methods. *Annales Zoologici Fennici* 47:46-56.

Oberhauser, K.
2004. Effects of female age, female mass, and nutrients from males on monarch eggs mass. pp. 21-26. *In* Oberhauser, K. S. & M. J. Solensky (Eds.), *The Monarch Butterfly: Biology and Conservation*. Cornell University Press, Ithaca, New York.

Oberhauser, K. & D. Frey
1999. Coercive mating by overwintering male monarch butterflies. pp. 67-78. *In* Hoth, J., L. Merino, K. Oberhauser, I. Pisanty, S. Price & T. Wilkinson (Eds.), *1997 North American Conference on the Monarch Butterfly*. Commission for Environmental Cooperation, Montreal, Quebec, Morelia, Michoacan, Mexico.

Oberhauser, K., R. Batalden & E. Howard
2009. *Monarch butterfly monitoring in North America. Overview of initiative and protocols*. Commission for Environmental Cooperation, Montreal, Canada. 55 pp.

Palmeirim, J. M.
1998. Analysis of skull measurements and measurers: Can we use data obtained by various observers? *Journal of Mammalogy* 79(3):1021-1028.

Palmer, A. & C. Strobeck
1986. Fluctuating asymmetry: Measurement, analysis, patterns. *Annual Review of Ecology and Systematics* 17:391-421.

PASW
2010. *PASW Statistics 18*. PASW, Chicago, Illinois.

Perini, T. A., G. L. De Oliverira, J. S. Ornellas & F. P. De Oliveira
2005. Technical error of measurement in anthropometry. *Revista Brasileira de Medicina do Esporte* 1(1):86-90.

Solensky, M. J. & K. S. Oberhauser
2004. Behavioral and genetic components of male mating success on monarchs. pp. 61-68. *In* Oberhauser, K. S. & M. J. Solensky (Eds.), *The Monarch Butterfly: Biology and Conservation*. Cornell University Press, Ithaca, New York.
2009. Male monarch butterflies, *Danaus plexippus*, adjust ejaculates in response to intensity of sperm competition. *Animal Behaviour* 77:465-472.

Soper, D.
2011. Post-hoc statistical power calculator for a Student t-test. Available at < *http://danielsoper.com/statcalc3/calc.aspx?id=49*> Last accessed 18 Sep 2011.

Stokes, M.
1985. Reliability and repeatability of methods for measuring muscle in physiotherapy. *Physiotherapy Theory and Practice* 1:71-76.

Tong, S., H. N. Cardinal, R. F. McLoughlin, D. B. Downey & A. Fenster
1998. Intra- and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging. *Ultrasound in Medicine and Biology.* 24(5):673-681.

Tuskes. P. M. & L. P. Brower
1978. Overwintering ecology of the monarch butterfly, *Danaus plexippus* L., in California. *Ecological Entomology* 3:141-153.

Ulijaszek , S. J. & D. A. Kerr
1999. Anthropometric error and the assessment of nutritional status. *British Journal of Nutrition* 82:165-177.

Van Hook, T.
1993. Non-random mating in monarch butterflies overwintering in Mexico. pp. 49-60. *In* Malcolm, S. B. & M. P. Zalucki, (Eds.), *Biology and Conservation of the Monarch Butterfly*. Natural History Museum of Los Angeles County, Los Angeles, California.
1996. *Monarch Butterfly Mating Ecology at a Mexican Overwintering Site: Proximate Causes of Non-Random Mating*. Ph.D. Dissertation, University of Florida, Gainesville, Florida, xiv + 242.

Williams, C. B.
1943. A safe method of measuring the wings of set butterflies. *Proceedings of the Royal Entomological Society of London (A)*18:3-4.

WHO Multicentre Growth Reference Study Group
2006. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. *Acta Paediatrica Suppl.* 450:38-46.

Yezerinac, S. M., S. C. Lougheed & P. Handford
1992. Measurement error and morphometric studies: Statistical power and observer experience. *Systematic Biology* 41(4):471-482.

York, H. A. & K. S. Oberhauser
2002. Effects of duration and timing of heat stress on monarch butterfly (*Danaus plexippus*) (Lepidoptera: Nymphalidae) development. *Journal of the Kansas Entomological Society* 75(4):290-298.

Yuan, X. P., D. Bach, A. Skanes & M. Drangova
2004. Assessment of intra- and interobserver variability of pulmonary vein measurement from CT angiography. *Academic Radiology* 11:1211-1218.

Zar, J. H.
1996. *Biostatistical Analysis*. 3rd ed. Prentice-Hall, Upper Saddle River, NJ, 662 pp.