LogNormal Distributions for Skin Area as a Function of Body Weight

David E. Burmaster
Alceon Corporation, PO Box 382669, Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954; deb@Alceon.com

Abstract

Using regression techniques, I re-analyzed the dataset cited by the US Environmental Protection Agency in its *Exposure Factors Handbook* that contains measurements of skin area, height, and body weight for 401 people spanning all stages of development. This re-analysis shows that a univariate model for total skin area as a function of body weight gives useful practical results with little or no loss of reliability as compared to the Agency's bivariate model. This new result leads to a new method to develop LogNormal distributions for total skin area as a function of body weight alone. Because these new methods replicate previously published results with much less effort, I recommend their use in probabilistic exposure assessments and risk assessments.

Introduction and Data

Dermal exposure to hazardous materials may result in risks to human health. To characterize this risk, the analyst needs to know the total skin area of the people exposed and the fraction of skin exposed.

Early in this century, Dubois and Dubois (1916) published a formula for predicting the total skin area as a function of the height and weight of a person. The form of the equation has proved useful and durable despite the fact that Dubois and Dubois had only nine observations:

$$SA = f(Ht, BW) = a\, Ht^{b}\, BW^{c} \qquad \text{Eqn 1}$$

where SA represents total skin area measured in m^2, Ht represents height measured in cm, BW represents body weight measured in kg, and a, b, and c are empirical constants found by fitting a plane to this equation transformed by the natural logarithm function:

$$\ln[SA] = \ln[a] + b \ln[Ht] + c \ln[BW] \qquad \text{Eqn 1a}$$

In the last 80 years, other researchers have made more measurements of total skin area and they have refined the fit of the equations (see especially: Boyd, 1935). In 1970, Gehan and George used 401 observations to revise the estimated values for parameters a, b, and c. Gehan and George estimated the parameters for three different age groups (<5 yr, 5 yr to 20 yr, and >20 yr), but they found that it was not necessary to estimate each group separately.

In 1985, the US Environmental Protection Agency (US EPA) published a report (Anderson et al., 1985) which re-analyzed the 401 observations used by Gehan and George. Using multiple regression (Draper & Smith, 1981) to fit the plane in Eqn 1a to the 401 observations, Anderson et al. found that $\hat{a} = 0.0239$, (or $\ln[\hat{a}] = -3.73$), $\hat{b} = 0.417$, and $\hat{c} = 0.517$ with $R^2 > 99$ percent. These parameters are all statistically significant since they have t-statistics with absolute values ranging from 7 to <195. These results agree with Gehan and George to two significant figures. Since then, the Agency has published these results in all editions of the *Exposure Factors Handbook* (US EPA, 1990; 1995; 1996), a document that influences most if not all of its regulatory programs involving exposure assessment.

In recent years, Murray and Burmaster (1992) and Phillips et al. (1993) have published methods to develop and/or simulate probability distributions for the variability of total skin area in a population. Since each of these methods has proven cumbersome for routine use in probabilistic exposure assessments, I undertook this research to seek a simple yet robust and reliable way to develop probability distributions for the total skin area in a population.

From various publications, principally Anderson et al. (1985) with confirmation from Gehan and George (1970) and Boyd (1935), I recreated the database relied upon by the US EPA (1990; 1995; 1996). This database includes measurements for 401 individuals: 161 people identified as males, 140 identified as females, and 100 not identified by gender. In this database, the males range in age from <1 mon to 794 mon [66 yr, 2 mon] and weigh from ~2 to ~80 kg; the females range in age from <1 mon to 456 mon [38 yr] and weigh from ~2 to ~98 kg; and those not identified by gender range in age from <1 mon to 600 mon [50 yr] and weigh from ~2 to ~92 kg. Overall, the database contains many children, few teens, and many adults. After not finding any additional data in more recent literature, I concur with the US Environmental Protection Agency's selection of these data as a good source of information about total skin area as a function of height and body weight.

Methods and Results

First, as a check, I refit Eqn 1a to the 401 observations using the multiple regression routines in Mathematica® (Wolfram, 1991). Table 1 shows the results in the upper left corner; they are identical to Anderson et al's (1985) results now widely disseminated in the US EPA's *Exposure Factors Handbook* (1990; 1995; 1996). I also fit Eqn 1a to the 161 observations for males and to the 140 observations for females. Table 1 shows these results in the top row. All the fitted parameters in the top row are statistically significant and each regression of the log-transformed variables has an $R^2 > 99$ percent.

As drawn by Mathematica® (Wickham-Jones, 1994), the 3D plots in the top panels in Figures 1, 2, and 3 show the data points and the best-fit plane for each group of people. (The best-fit plane is displaced by -0.5 so we can see all the data points.) The 2D plots in the bottom panels Figures 1, 2, and 3 show the residuals from the plane. While the residuals in the lower panels are acceptably small and show no apparent trend with increasing body weight, the strong co-linearity of the data points visible in the upper panels is a warning sign that multiple regression may not be appropriate (Draper & Smith, 1981). Additional plots and analyses of the log-transformed data show that regression of ln[BW] against ln[Ht] has an $R^2 \geq 97.7$ percent for all three groups. The information contained in Ht and BW is nearly redundant.

Table 1 shows the results of the fitting two simplified models to the data for each of the three groups. With c = 0, Eqn 1 becomes:

$$SA = g_1(\, Ht\,) = a\, Ht^{\,b} \qquad \text{Eqn 2}$$

and with b = 0, Eqn 1 becomes:

$$SA = g_2(\, BW\,) = a\, BW^{\,c} \qquad \text{Eqn 3}$$

Based on linear regressions of the log-transformations of Eqns 2 and 3, the center and bottom rows of Table 1 show the best-fit parameters. From visual inspection of the fitted lines and of the residuals from the fits, and from the consistent pattern in the R^2 values for the regressions, I conclude that Eqn 3 outperforms Eqn 2 as a predictor of total skin area. The top panels in Figures 4, 5, and 6 show the best-fit straight line for Eqn 3 and the center panels in the same figures show the residuals from that line. Based on the visual comparison of the residuals from the log-transformations of Eqns 1 and 3 to a 45 degree line as shown in the lower panels, I conclude that Eqn 3 does as well as Eqn 1 in predicting total skin area, but Eqn 3 has the advantage of having only one explanatory variable. In practice, this is a big advantage with little or no loss of accuracy.

Brainard and Burmaster (1992) and Burmaster and Crouch (1997) have shown that the body weights of males and females follow LogNormal distributions of this form (Evans et al, 1993; Burmaster & Hull, 1997):

$$BW \sim \exp[\, Normal[\, \mu_{BW}, \sigma_{BW} \,] \,] \qquad \text{Eqn 4}$$

which is equivalent to:

$$\ln[\, BW \,] \sim Normal[\, \mu_{BW}, \sigma_{BW} \,] \qquad \text{Eqn 4a}$$

Here, exp[•] represents the exponential function, ln[•] represents the Napierian (or natural) logarithm function, and Normal[μ, σ] represents the Normal or Gaussian distribution with mean μ and standard deviation σ (with $\sigma > 0$).

When Eqns 3 and 4 obtain, total skin area follows a LogNormal distribution that is a function of body weight:

$$SA \sim a \bullet (\, \exp[\, Normal[\, \mu_{BW}, \sigma_{BW} \,] \,] \,)^{c} \qquad \text{Eqn 5}$$

$$\sim \exp[\, Normal[\, c \bullet \mu_{BW} + \ln[\, a \,], c \bullet \sigma_{BW} \,] \,] \,;\, c > 0 \qquad \text{Eqn 5a}$$

Thus, the distribution for SA has this LogNormal form:

$$SA \sim \exp[\, Normal[\, \mu_{SA}, \sigma_{SA} \,] \,] \qquad \text{Eqn 6}$$

with:

$$\mu_{SA} = c \bullet \mu_{BW} + \ln[\, a \,] \qquad \text{Eqn 7}$$

$$\sigma_{SA} = c \bullet \sigma_{BW} \qquad \text{Eqn 8}$$

When $\ln[\, \hat{a} \,] = -2.2781$ and $\hat{c} = 0.6821$ as estimated for the 401 people in the database,

$$\mu_{SA} = 0.6821 \bullet \mu_{BW} - 2.2781 \qquad \text{Eqn 9}$$

$$\sigma_{SA} = 0.6821 \bullet \sigma_{BW} \qquad \text{Eqn 10}$$

The results from Eqns 9 and 10 (for all ages) in combination with the results from Table 3 (for adults) in Brainard and Burmaster (1992) reconfirm the results (for adults) in Murray and Burmaster (1992) to within a tight tolerance (taking care to convert the units correctly). As an example, on page 459, Murray and Burmaster recommend this LogNormal distribution for the skin area measured in m^2 for adult men:

$$SA_{men} \sim \exp[\text{ Normal}[\ 0.671,\ 0.096\ ]\ ] \qquad \text{Eqn 11}$$

while Eqns 9 and 10 and the results for adult men from Table 3 in Brainard and Burmaster predict this LogNormal distribution in the same units:

$$SA_{men} \sim \exp[\text{ Normal}[\ 0.689,\ 0.116]\ ] \qquad \text{Eqn 12}$$

Above its median, Eqn 12 overpredicts Eqn 11 by a small amount because its parameters are each numerically larger. For example, at two geometric standard deviations above the geometric mean, Eqn 12 overpredicts Eqn 11 by ~5 percent.

Dividing each side of Eqn 3 by BW yields a new relationship that may be useful in risk assessments involving only the dermal pathway:

$$\frac{SA}{BW} = a\ BW^{\,c-1} \qquad \text{Eqn 13}$$

Since $0 < \hat{c} < 1$, Eqn 14, next, needs selective use of the absolute value function (denoted |x|) for the second parameter only of the LogNormal distribution) (Keeping, 1995):

$$\frac{SA}{BW} \sim \exp[\text{ Normal}[\ (c - 1) \bullet \mu_{BW} + \ln[\ a\ ],\ |c - 1| \bullet \sigma_{BW}\ ]\ ] \qquad \text{Eqn 14}$$

Because SA and BW are perfectly correlated ($\rho = 1$) in this model, the variance for ratio of the SA/BW is smaller than the variance of BW alone.

Conclusions and Discussion

For all practical purposes, the log-transformation of Eqn 3 provides as good a fit to each of the three datasets (for 401 people, for 161 males, and for 140 females) as does the log-transformation of Eqn 1. The resulting estimates of $\hat{c}$ in Eqn 3 for scaling skin areas as a function of body weight for people of all ages compares well with $\hat{c} = 0.69$ published earlier (Boyd, 1935; see also discussions for other species in Calder, 1996). The value found here compares well to the theoretical value (2/3) for spheres, cubes, tetrahedra, or any family of self-similar solids (Thompson, 1961; Rhomberg, 1997).

This first result -- in combination with the results for adults from Brainard and Burmaster (1992) and with the results for all ages from birth to age 74 years from Burmaster and Crouch (1997) -- yields LogNormal distributions for total skin area as a function of body weight. Based on the results in these three papers, I recommend this new method (Eqns 6, 7, 8, 9, and 10) as equally reliable and much less cumbersome than the methods previously recommended in Murray and Burmaster (1992) and Phillips et al. (1993).

When used with great care to avoid the severe problems and spurious results caused by failing to consider correlations and instantiations in probabilistic risk assessments, Eqns 13 and 14 for the distribution of the ratio SA/BW may have some limited usefulness in risk assessments involving only the dermal pathway. However, as soon as a risk assessment involves additional pathways, Eqns 13 and 14 may lead to large numerical errors as discussed cogently in Smith et al. (1992) and Ferson (1996). In calculations involving multiple pathways, not just the dermal pathway, it is important to avoid the use of Eqn 14 since there is no way to preserve the proper correlations with it.

Dedication

I dedicate this paper to George Wald.

Acknowledgments

I thank Lorenz R. Rhomberg and three anonymous reviewers for excellent suggestions to improve this manuscript. Alceon Corporation funded this research.

Trademarks

Mathematica ® is a registered trademark of Wolfram Research, Inc: http://www.wri.com
Alceon ® is a registered trademark of Alceon Corporation: http://www.alceon.com

References

Anderson et al, 1985
Anderson, E., N. Browne, S. Duletsky, J. Ramig, and T. Warn, 1985, Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments. Office of Health and Environmental Assessment. US EPA/600/8-85/010, Washington, DC

Boyd, 1935
Boyd, E., 1935, The Growth of Surface Area of the Human Body, University of Minnesota, Minneapolis, MN

Brainard & Burmaster, 1992
Brainard, J. and D.E. Burmaster, 1992, Bivariate Distributions for Height and Weight of Men and Women in the United States, Risk Analysis, 1992, Volume 12, Number 2, pp 267 - 275

Burmaster & Crouch, 1997
Burmaster, D.E. and E.A.C. Crouch, 1997, Lognormal Distributions for Body Weight as a Function of Age for Males and Females in the United States, 1976 - 1980, Risk Analysis, in press

Burmaster & Hull, 1997
Burmaster, D.E. and D.A. Hull, 1997, Using LogNormal Distributions and LogNormal Probability Plots in Probabilistic Risk Assessment, Human and Ecological Risk Assessment, in press

Calder, 1996
Calder, III, W.A., 1996, Size, Function, and Life History, Dover Publications, Mineola, NY

Draper & Smith, 1981
Draper, N.R. and H. Smith, 1981, Applied Regression Analysis, Second Edition, John Wiley & Sons, New York, NY

Dubois & Dubois, 1916
Dubois, D. and E.F. Dubois, 1916, A Formula to Estimate the Approximate Surface Area If Height and Weight Be Known, Archives of International Medicine, Volume 17, pp 863 - 871

Evans et al, 1993
Evans, M., N. Hastings, and B. Peacock, 1993, Statistical Distributions, Second Edition, John Wiley & Sons, New York, NY

Ferson, 1996
Ferson, S., 1996, Automated Quality Assurance Checks on Model Structure in Ecological Risk Assessments, Human and Ecological Risk Assessment, Volume 2, Number 3, pp 558 - 569

Gehan & George, 1970
Gehan, E.A. and S.L. George, 1970, Estimation of Human Body Surface Area from Height and Weight, Cancer Chemotherapy Reports, Volume 1, Number 4, pp 225 - 235

Keeping, 1995
Keeping, E.S., 1995, Introduction to Statistical Inference, Dover, New York, NY

Murray & Burmaster, 1992
Murray, D.M., and D.E. Burmaster, 1992, Estimated Distributions for Total Body Surface Area of Men and Women in the United States, Journal of Exposure Analysis and Environmental Epidemiology Volume 2, Number 4, pp 451 - 461

Phillips et al, 1993
Phillips, L.J., R.J. Fares, and L.G. Schweer, 1993, Distributions of Total Skin Surface Area to Body Weight Ratios for Use in Dermal Exposure Assessments, Journal of Exposure Analysis and Environmental Epidemiology, Volume 3, Number 3, pp 331 - 338

Rhomberg, 1997
Rhomberg., L.R., 1997, personal communication via email.

Smith et al, 1992
Smith, A.E., P.B. Ryan, and J.S. Evans, 1992, The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating Population Distribution of Risk, Risk Analysis, Volume 12, Number 4, pp 467 - 474, December 1992

Thompson, 1961
Thompson, D'A., 1961, On Growth and Form, Abridged and edited by J.T. Bonner, Cambridge University Press, Cambridge, UK

US EPA, 1990
US Environmental Protection Agency, 1990, Exposure Factors Handbook, Office of Health and Environmental Assessment, EPA/600/8-89/043, March 1990

US EPA, 1995
US Environmental Protection Agency, 1995, Exposure Factors Handbook, External Review Draft, Exposure Assessment Group, EPA/600/P-95/002A, Washington, DC, June 1995

US EPA, 1996
US Environmental Protection Agency, 1996, Exposure Factors Handbook, Science Advisory Board Review Draft, Exposure Assessment Group, EPA/600/P-96/002B a, b, and c, Washington, DC, August 1996

Wickham-Jones, 1994
Wickham-Jones, T., 1994, Mathematica Graphics, Techniques & Applications, Springer-Verlag, Telos, Santa Clara, CA

Wolfram, 1991
Wolfram, S., 1991, Mathematica®, A System for Doing Mathematics by Computer, Second Edition, Addison- Wesley, Redwood City, CA

deb 14 June 97
proofed

Table 1
Results from the Regressions
for Skin Area as a Function of Height and/or Body Weight

	• For 401 People --- >>				• For 161 Males --- >>				• For 140 Females --- >>			
	ln(ahat) •••••	bhat •••••	chat •••••	adjR2 •••••	ln(ahat) •••••	bhat •••••	chat •••••	adjR2 •••••	ln(ahat) •••••	bhat •••••	chat •••••	adjR2 •••••
Eqn 1: SA = f(Ht, BW)	-3.7330	0.4170	0.5170	0.9921	-3.5933	0.3771	0.5371	0.9937	-3.3909	0.3209	0.5496	0.9961
Eqn 2: SA = g1(Ht)	-8.1700	1.6963	zero	0.9806	-8.1784	1.6984	zero	0.9798	-8.2014	1.7019	zero	0.9860
Eqn 3: SA = g2(BW)	-2.2781	zero	0.6821	0.9909	-2.2752	zero	0.6868	0.9926	-2.2678	zero	0.6754	0.9956

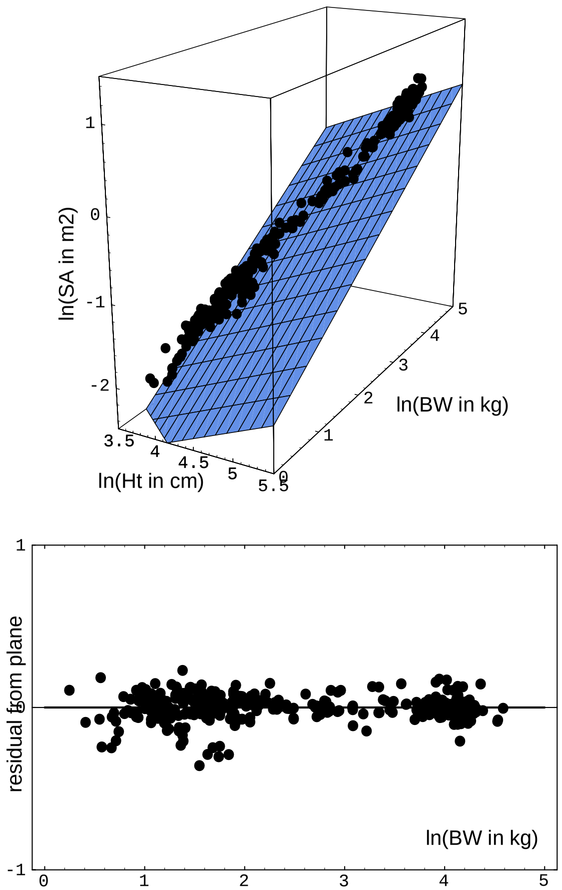

Figure 1 for
401 People

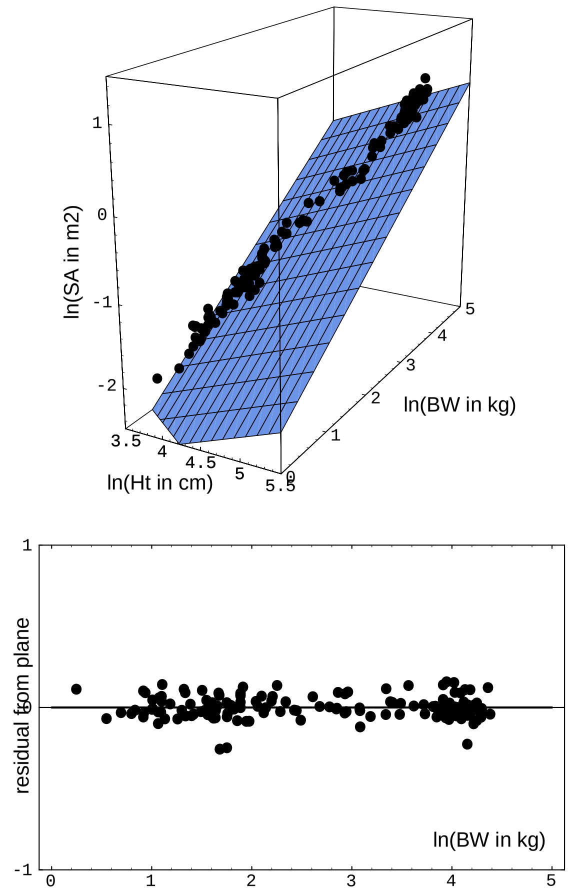

Figure 2 for
161 Males

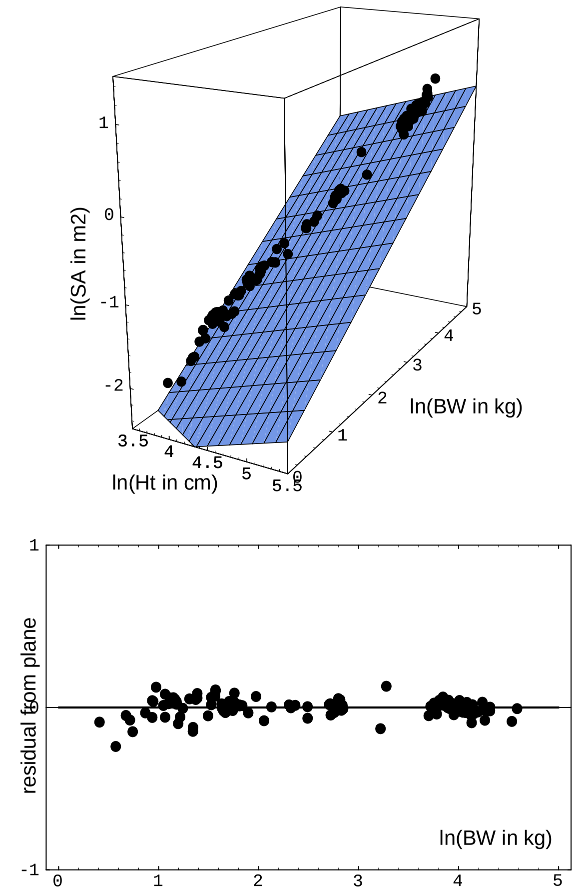

Figure 3 for
140 Females

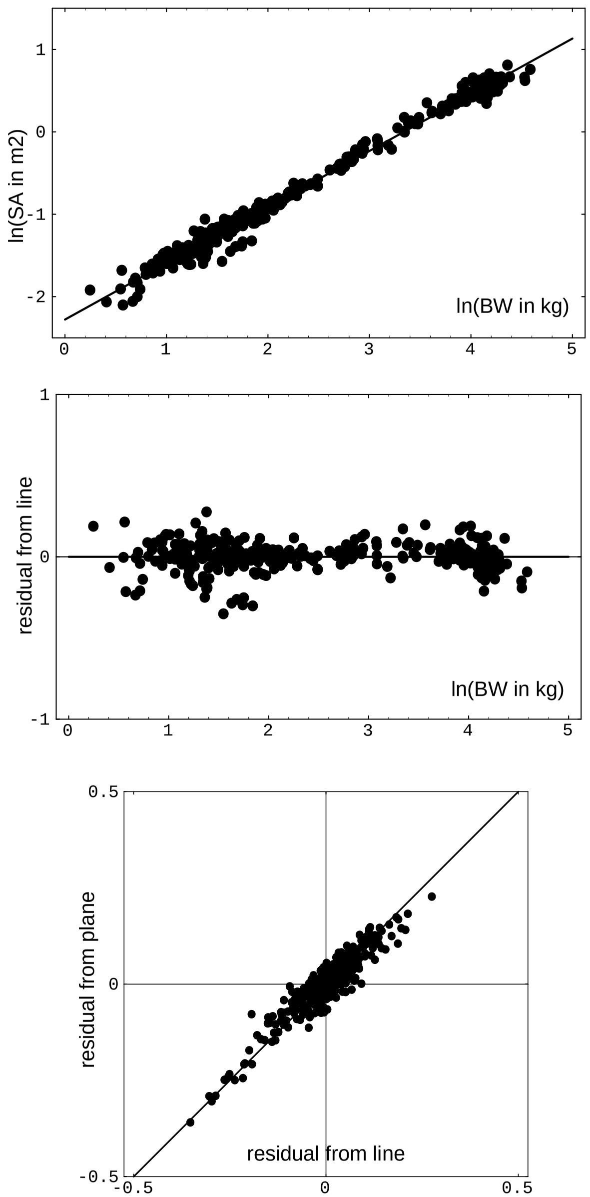

Figure 4 for
401 People

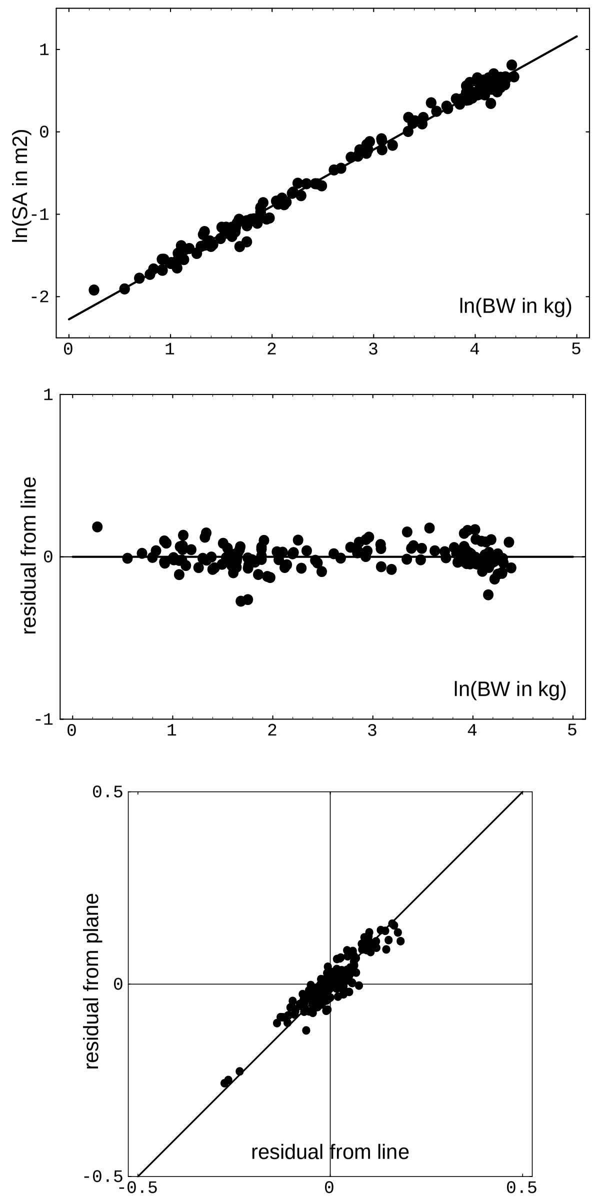

Figure 5 for
161 Males

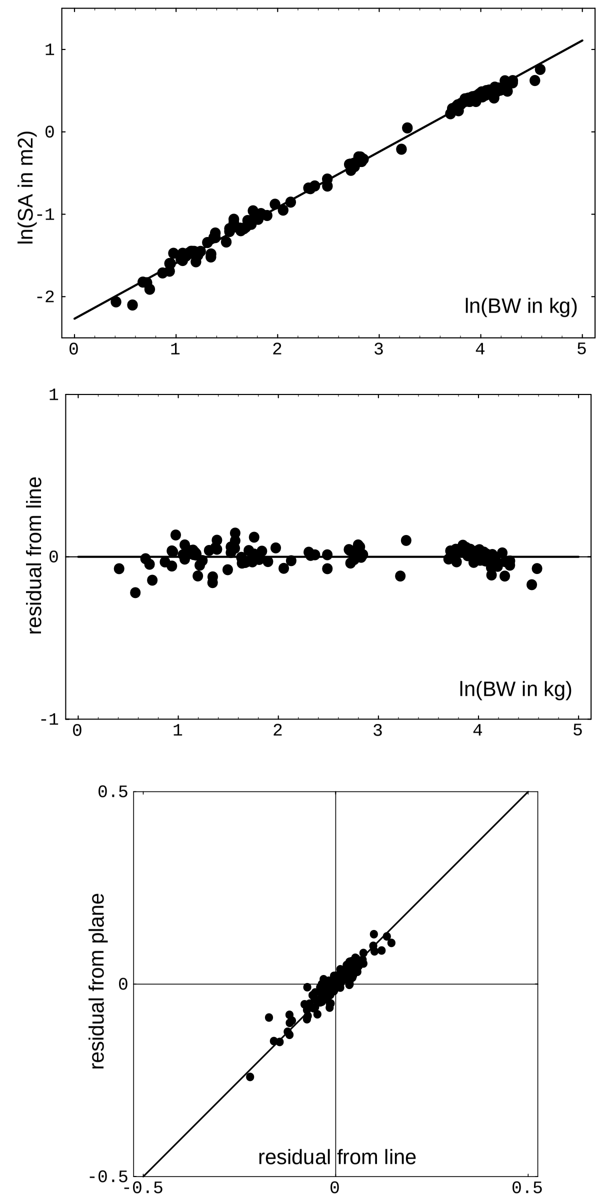

Figure 6 for
140 Females