Claire C. Gordon,[1] *Ph.D. and Jane E. Buikstra,*[2] *Ph.D.*

Linear Models for the Prediction of Stature from Foot and Boot Dimensions

REFERENCE: Gordon, C. C. and Buikstra, J. E., "**Linear Models for the Prediction of Stature from Foot and Boot Dimensions,**" *Journal of Forensic Sciences*, JFSCA, Vol. 37, No. 3, May 1992, pp. 771–782.

ABSTRACT: Estimation of stature from the dimensions of foot or shoeprints has considerable forensic value in developing descriptions of suspects from evidence at the crime scene and in corroborating height estimates from witnesses. This study extends the findings of previous researchers by exploring linear models with and without gender and race indicators, and by validating the most promising models on a large, recently collected military database. Boot size and outsole dimensions are also examined as predictors of stature.

The results of this study indicate that models containing both foot length and foot breadth are significantly better than those containing only foot length. Models with race/gender indicators also perform significantly better than do models without race/gender indicators. However, the difference in performance is slight, and the availability of reliable gender and race information in most forensic situations is uncertain. Analogous results were obtained for models utilizing boot size/width and outsole length/width, and in this study these variables performed nearly as well as the foot dimensions themselves.

Although the adjusted R^2 values for these models clearly reflect a strong relationship between foot/boot length and stature, individual 95% prediction limits for even the best models are ±86 mm (3.4 in.). This suggests that models estimating stature from foot/shoeprints may be useful in the development of subject descriptions early in a case but, because of their imprecision, may not always be helpful in excluding individual suspects from consideration.

KEYWORDS: physical anthropology, anthropometry, height estimation, footprints

Estimation of stature from foot or shoeprints has considerable forensic science value in developing descriptions of suspects from evidence at the crime scene and in corroborating height estimates from witnesses. Historically, two approaches to this problem have been taken: expression of foot length as a percentage of stature, and least-squares regression with stature as the dependent variable and foot length as the independent variable [*1,2*]. Recent work by Giles and Vallandigham [*2*] suggests that regression is preferable since the variance of its predictive errors is lower than that for the percentage estimation technique. Linear models also permit the inclusion of gender and racial indicator variables which may substantially improve stature predictions.

This study extends the findings of previous researchers by examining a variety of linear models to estimate stature from foot dimensions. The most complete study in this area

Received for publication 25 July 1991; accepted for publication 8 Oct. 1991.

[1] Senior anthropologist, Science & Advanced Technology Directorate, U.S. Army Natick Research, Development and Engineering Center, Natick, MA.

[2] Harold H. Swift Distinguished Service Professor, Department of Anthropology, University of Chicago, Chicago, IL.

to date, Giles and Vallandigham [2], treated genders separately and pooled races. In contrast, this study compares the performance of models with and without gender and race indicator variables, and the most promising models are independently validated on a large, recently collected military database. Validation included close examination to determine whether equations lacking gender/race specificity penalize any particular demographic group in their accuracy. In addition, the efficacy of boot size and outsole dimensions as predictors of stature is also examined.

Materials

Two independently assembled U.S. Army databases are used in this study. Data from a relatively small study ($n = 867$) with an extensive set of foot measurements and which included professional boot-fitting are used to develop candidate models for stature estimation. The best of these models are then tested on a much larger database ($n = 3982$) which lacks data on boot size but includes a wider representation of ages and races.

Modeling Data Base

Data used in model building come from the 1985 Combat Boot Fit Test conducted at Fort Jackson, South Carolina [*3*]. A total of 867 soldiers participated in this test. Two thirds of these were in Advanced Individual Training in administrative job categories, the remaining subjects consisted of instructors and medical personnel. Each subject was measured for 31 different foot and lower leg dimensions along with height and weight. Supporting biographical data, including age, gender, and race were self-reported by the subjects. Race was categorized as White, Black, Asian/Pacific Islander, and American Indian. In addition, each subject was also fit with a standard Army combat boot, which comes in some 111 sizes and widths, and their best-fitting boot size and width was recorded.

Anthropometric data in this study were collected by a team of six measurers specifically recruited and trained for this study. Measurers were paired, assigned to one of three measuring stations, and learned only the subset of dimensions at their station. One measurer at each station recorded while the other measured, and they switched at will to alleviate fatigue and boredom. Thus only two individuals contributed to observer error for any dimension in the study. These methods resulted in relatively low mean absolute differences and relatively high reliability coefficients for stature (5.0 mm; 99.8%), foot length (1.2 mm; 99.5%), and foot breadth (2.5 mm; 88.0%). And unlike earlier Army databases, such as those used by Giles and Vallandigham [2], male and female subjects were measured with exactly the same techniques, by the same measurers. Boot fitting in the 1985 study was undertaken by two full-time professional fitters borrowed from the Fort Jackson Central Initial Issue Facility. Data on combat boot outsole dimensions were taken from Army specifications and confirmed on a set of master patterns.

Of the 867 participants in the 1985 study, 836 were either White or Black, and so initial modeling efforts focused on these individuals. Their race and gender distributions are given in Table 1. Their ages ranged between 17 and 56 years, with a median of 20 years. Stature in the group varied between 1415 and 1942 mm. Full descriptive statistics for stature, foot length, and foot breadth are given in Table 2.

Validation Database

Because the median age of the modeling database subjects was so young, and because only Blacks and Whites were modeled initially, a critical aspect of this study involved model validation on an independent database with a broader racial distribution and better

TABLE 1—*Race and gender distributions in the modeling sample.*

	Females	Males
White	278	199
Black	279	80

TABLE 2—*Descriptive statistics for the modeling sample.*

	Stature, mm		Foot Length, mm		Foot Breadth, mm	
	Females	Males	Females	Males	Females	Males
Mean	1619.4	1757.1	244.9	269.8	91.6	100.9
S.d.[a]	65.9	71.5	13.2	13.4	4.6	5.4
Min	1414	1564	193	237	77	87
Max	1809	1942	298	303	107	118
n	574	293	574	292	573	291

[a] Standard deviation.
NOTE: 1 mm = 0.039 in.

representation of mature adults than provided by earlier Army databases. The data used for model validation in this study came from the 1988 U.S. Army Anthropometric Survey [*4*]. Some 2208 females and 1774 males from the 1988 survey are used in this study.

Like the 1985 combat boot study, the 1988 survey collected self-reported racial information. However, the 1988 survey employed a slightly different racial categorization in that Whites and Blacks were identified specifically as *not* of Hispanic origin, and a separate Hispanic racial/ethnic category was provided. Self-reported racial data were verified on site in the 1988 survey by a single professional anthropologist who interviewed all subjects. The addition of a separate Hispanic category in 1988 reflected a need to insure that sufficient numbers of Hispanic soldiers were sampled in order to address any unique aspects of their anthropometry that might impact protective clothing and equipment design.

The racial composition of the validation sample is given in Table 3, and exactly matches that of the contemporary active-duty Army. It is considerably more diverse than that of the 1985 study, and provides an opportunity to test models without a race indicator on subjects not of Black or White extraction. In addition to a more diverse racial composition, the 1988 validation database has considerably more mature adults in it. The males range between 17 and 51 years of age with a median age of 25 years; the females range between 18 and 50 years of age with a median age of 24.5 years. In addition, subjects in the 1988 study sample came from a wider variety of military occupations than the 1985 study since the 1988 sampling strategy intentionally selected subjects from a cross section of military occupational specialties [*4*], whereas the 1985 study did not. Stature also has a wider range in the 1988 study sample: 1428 to 2042 mm. Full descriptive statistics for stature, foot length, and foot breadth are given in Table 4, and a glance at these indicates that the distributions in the 1988 survey are very close to those in the 1985 study, despite the demographic differences in the two samples.

Combat boots were not fitted to subjects as part of the 1988 survey, and so stature estimation models based upon boot size or outsole dimensions or both could not be validated. Protocols for measuring stature, foot length, and foot breadth were identical in the 1985 and 1988 studies. As in the 1985 study, a maximum of two measurers contributed to the observer error in any dimension. Data reliability was also high in the

TABLE 3—*Race and gender distributions in the validation sample.*

	Females	Males
White	1140	1172
Black	922	458
Hispanic	58	68
Asian/Pacific	32	28
Native American	14	12
Mixed/other	42	36

TABLE 4—*Descriptive statistics for the validation sample.*

	Stature, mm		Foot Length, mm		Foot Breadth, mm	
	Females	Males	Females	Males	Females	Males
Mean	1629.4	1755.8	244.4	269.7	89.7	100.6
S.d.[a]	63.6	66.8	12.2	13.1	4.9	5.3
Min	1428	1497	203	228	73	80
Max	1870	2042	290	310	109	122
n	2208	1774	2208	1774	2208	1774

[a] Standard deviation.

Army's 1988 survey; mean absolute differences and reliability coefficients were as follows: stature 3.23 mm and 99.9%; foot length 0.74 mm and 99.8%; foot breadth 0.88 mm and 98.5%.

Methods

Preliminary Analyses

Both modeling and validation databases were edited before their use in this study. The editing routines used are described in detail elsewhere [5]; they use three primary methods of identifying outliers: range checks, regression estimates, and extreme value review. In the case of the 1985 modeling database, the data were originally recorded by hand, keypunched afterward, and then edited. Bad values located in the editing process were declared missing; however, this affected only four values in this study.

In the case of the 1988 validation database, the data were entered and edited on-site, so the vast majority of outliers were identified and either validated or corrected immediately by remeasuring the subject. The 1988 database was also subjected to postsurvey editing routines, but these identified only an additional 0.02% bad values [6]. Many bad values were due to electronic media failure, and these were restored from printed copies of the original data; the others were declared missing. None of the missing values in the 1988 survey affected the sample in this study.

In addition to data cleaning, a number of exploratory analyses were conducted to ensure that the data in this study met normality and homoscedasticity assumptions for least-squares regression analysis. Stem and leaf plots, box plots, and spread versus level plots [7] were examined for the modeling database as a whole and for each gender/race subgroup. Nothing in these preliminary analyses suggested the need for data transformation or for robust or nonparametric statistical methods, or both, and so they are not discussed further.

Model Building

Least-squares regression [*8*] was the method chosen for model building. Two foot dimensions and two demographic variables were selected for evaluation in a stature (ST) prediction model: foot length (FL), ball of foot breadth horizontal (FB), race, and gender. Other foot dimensions were considered, including heel breadth, ball of foot breadth, and ball of foot length. However, none of these can be easily derived from footprint data, and a preliminary review of simple and multiple correlation coefficients indicated that foot length and foot breadth horizontal were the strongest candidates for stature estimation. Race and gender were also chosen for model building since these two variables are known to influence allometric variation in so many body dimensions, because foot length/stature ratios appear to vary among race/gender groups [*9*], and because race and gender have not been addressed analytically within foot/stature predictive models to date. Race was coded 0 for Whites and 1 for Blacks; gender was coded 0 for males and 1 for females.

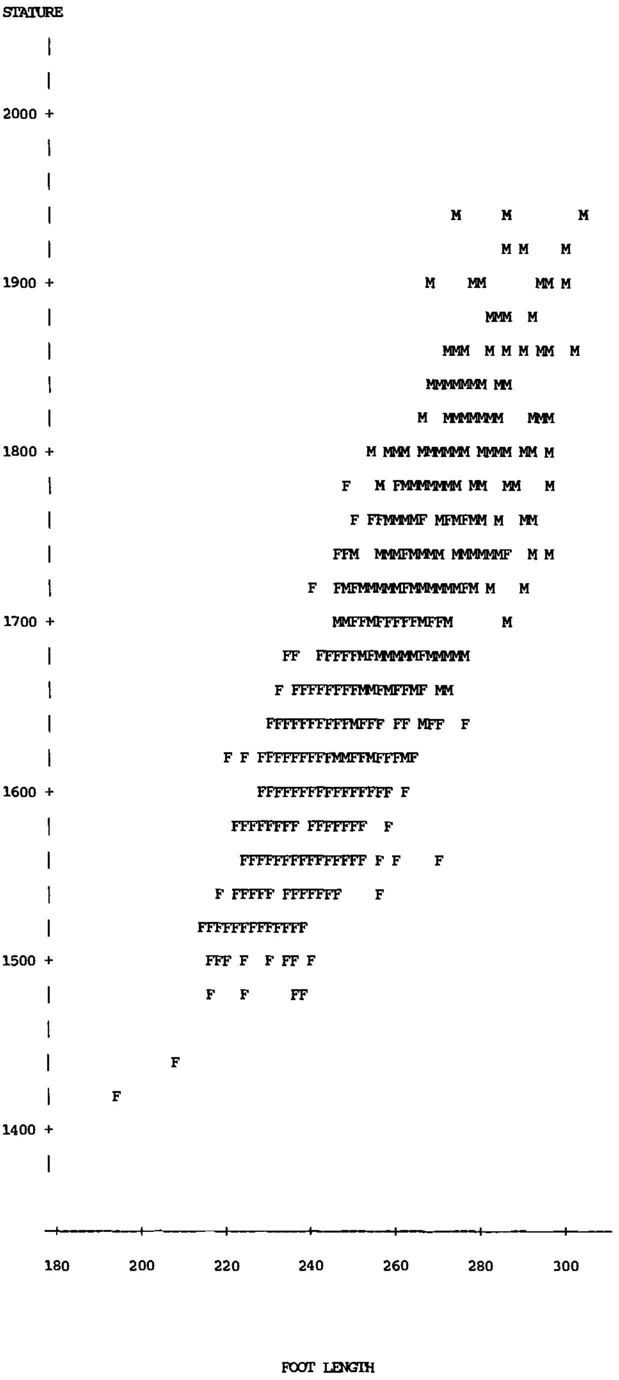
Note that most other researchers start with the assumption that males and females should be modeled separately. Preliminary analyses in this study suggested that the male and female slopes for an FL/ST regression were not very different from each other (3.78 for males; 3.72 for females). And as can be seen in Fig. 1, there is good reason to believe that a single linear model could adequately describe both groups. Thus in this study we start with a mixed gender database and address gender differences by including indicator variables in the linear models. Eight models to estimate stature from foot dimensions were generated: five when the race or gender of the suspect or both are known, and three when the race/gender are unknown.

Several types of residual diagnostics were undertaken for the estimated models to determine whether each model adequately fit the data. Raw residuals were plotted against both foot length and predicted stature, and these plots were examined for any patterns suggestive of the need for a different type of regression or for higher-order terms. These residual plots included racial/gender subgroup coding to see if any particular demographic subgroup was poorly fit by the model under study. In addition, all residuals larger in magnitude than twice their standard errors were output and examined individually to determine whether any demographic subgroup or any tails of the foot or stature distributions were particularly disadvantaged in the model under study. None of the models estimated in this study exhibited anything unusual in their residuals, and so residual diagnostics are not discussed further.

Because several models may adequately fit the same data set, some criteria are needed to identify the optimal model. In this study, four approaches were taken in evaluating model performance: partial F-tests between full and reduced models, adjusted R^2 values, width of 95% prediction intervals, and performance on the validation database.

Partial F-tests were used in a stepdown approach to determine whether the variable(s) deleted in each reduced model had added significant explanatory power to the prior model, given the presence of variables common to both [*8*]. This approach helped to identify the most parsimonious model that could adequately fit the data. Adjusted R^2 values were also calculated to indicate the percentage of variation in stature that was explained by each regression, adjusted for the number of independent variables in the model [*10*]. In general, the larger the adjusted R^2 value, the better the model is at explaining variation in stature.

Whereas adjusted R^2 magnitudes are useful in comparing the percentage of stature variation that competing models can explain, they do not indicate the precision with which one can estimate stature from a particular model. One way to express the precision of a stature estimate is the width of a 95% confidence interval for that estimate. We refer to these confidence intervals for individual estimates as "prediction intervals" to avoid confusing them with confidence intervals for estimates of the dependent variable's



---

* 512 obs hidden.

FIG. 1—*Plot of stature and foot length by gender.**

mean [8]. And as has been pointed out previously by Giles and Klepinger [11], properly constructed prediction intervals are particularly relevant in forensic science applications because the narrower the interval is, the more specific the suspect description and expert testimony can be. In this study, SAS software was used to generate 95% confidence limits for stature predictions at each point in the data set. As would be expected with prediction limits, the width of the 95% prediction interval varied as a function of both distance from the FL and FB means and gender/race group membership. However, the magnitude of variation in width of 95% prediction intervals was quite small for all models, and so the minimum and maximum widths are reported with each model studied.

Partial *F*-tests, prediction interval widths, and adjusted R^2 values were used to select several recommended models for further examination. These models were then applied to the validation database, an adjusted R^2 was calculated, and residuals were examined to assess performance over the 1988 database and within its demographic subgroups.

Results

Estimation of Stature from Foot Data: Model Building

The results of step-down model building for the prediction of stature from foot dimensions are presented in Table 5. The full model, which included terms for FL, FB, race, gender, and their nine possible interactions, explained 78.4% of the variation in stature for the 1985 sample. All interaction terms were removed in the next step, and a partial *F*-test of the main effects model against the full model was conducted. The *F*-statistic was not significant at the 0.05 level, and so the null hypothesis that coefficients for all the interaction terms are 0 was not rejected, and the main effects model was accepted. The fact that no interaction terms are required in this model indicates that the slopes for FL and FB are essentially the same in the four race/gender subgroups.

In the next step, FB was removed and a partial *F*-test of the reduced model against the main effects model was conducted to verify that FB was actually needed. As can be seen in Table 5, this F test was significant at the 0.05 level, and so we reject the hypothesis that the coefficient for FB equals 0, and retain FB in the model. In the next two steps, race and gender were each eliminated in turn and the reduced models were tested against the main effects model with significant results in each case.

We thus conclude that the main effects model with FL, FB, race, and gender is the most parsimonious model that adequately fits these data. This form of the model assumes parallel FL/FB-STAT relationships in the four demographic subgroups, but provides separate intercepts for each group. Each group's intercept is specified as the sum of the common intercept term and the appropriate race/gender terms.

Because one may not always have supplementary information on the race and gender

TABLE 5—*Model building based on foot data.*

Model	Adjusted R^2	Test	F (df)
1. Full model	0.784	...	...
2. FL FB race gender	0.784	2 vs. 1	0.70 (9817)
3. FL race gender	0.781	3 vs. 2	5.47 (1817)[a]
4. FL FB race	0.779	4 vs. 2	21.26 (1817)[b]
5. FL FB gender	0.754	5 vs. 2	115.52 (1817)[b]
6. FL FB FLFB	0.734	...	...
7. FL FB	0.735	7 vs. 6	0.33 (1827)
8. FL	0.725	8 vs. 7	30.70 (1828)[b]

[a] $P \leq 0.05$.
[b] $P \leq 0.001$.

of a suspect, models using footprint data alone were also examined (see Table 5). In this case, both FL and FB are required, but no interaction terms are necessary. Unlike the previous model, this form does not permit separate intercepts for each demographic subgroup, with some loss of explanatory power as a result (78.4 − 73.5% = 4.9%).

The parameter estimates for each of the final models are presented in Table 6, along with their minimum and maximum 95% prediction interval widths. As can be seen in Tables 5 and 6, although the model with race and gender is statistically superior to the model with foot dimensions only, their prediction interval widths differ by a maximum of only ±10 mm (0.4 in.), which is not a large figure when the best predictions obtainable are of the order ±86 mm (3.4 in.).

Model Validation

Tables 7 and 8 present the results of model validation on the 1988 database. Predicted statures were calculated for each of the subjects in the 1988 database using Models 2 and 7. These were subtracted from the individuals' actual statures, and the residuals were used to calculate adjusted R^2 statistics. Because both models were derived on a database composed solely of Black and White subjects (which included Hispanics pooled with Whites), initial validations focused on the Black/White/Hispanic subset of the 1988 study to determine whether the regressions were performing adequately in a comparable, but independently sampled, population. Then each of the equations was validated separately against each of the racial/ethnic groups in the 1988 study: Whites, Blacks, Hispanics, Asian/Pacific Islanders, and Native Americans.

Table 7 reports validation results for Model 2: FL FB race gender. This model was the optimal model for the 1985 modeling database. The adjusted R^2 value of 74% for this model on Whites and Blacks in the 1988 validation database compares favorably with that based on the 1985 modeling data (78%). The mean residual for Model 2 applied to the 1988 study is also promising: 6.67 mm. The fact that we have a positive mean indicates that Model 2 slightly underestimates stature in the 1988 database. This same

TABLE 6—*Parameter estimates for recommended models based on foot data.*

Model	FL	FB	Race	Gender	Intercept	95% PI[a] Widths
2. FL FB race gender	4.04	0.75	−35.73	−22.80	601.13	(±85.7, ±86.5)
7. FL FB	3.97	2.08	. . .	. . .	462.00	(±95.0, ±95.7)

[a] Prediction interval.

TABLE 7—*Validation of Model 2: FL FB race gender.*

		Residuals	
Demographic Group	Adjusted R^2	Mean, mm	Standard Deviation, mm
(Whites, Hispanics, Blacks)	0.7438	6.67	45.08
Whites	0.7358	8.80	45.56
Blacks	0.7507	6.28	42.99
Hispanics[a]	0.6302	−26.08	44.00
Asian/Pacific Islanders[a]	0.7226	−21.82	40.59
Native Americans[a]	0.7651	−0.61	43.52
Hispanics[b]	0.7099	7.55	44.42
Asian/Pacific Islanders[b]	0.7571	14.84	41.06

[a] Coded as Whites.
[b] Coded as Blacks.

TABLE 8—*Validation of Model 7: FL FB.*

		Residuals	
Demographic Group	Adj. R^2	Mean, mm	Standard Deviation, mm
(Whites, Hispanics, Blacks)	0.6834	12.39	49.11
Whites	0.6444	28.83	45.47
Blacks	0.7299	−13.32	43.21
Hispanics	0.6986	−5.53	45.88
Asian/Pacific Islanders	0.7690	−2.88	42.03
Native Americans	0.7046	−16.38	45.90

phenomenon was noted by Giles and Vallandigham [2] in their validation of a regression based on a relatively young military sample. As was the case with their samples, the slight underestimation of stature in the validation sample is probably due to the fact that a larger proportion of individuals in the validation database have reached their full adult height.

Table 7 also presents separate tests of the FL FB race gender model (Model 2) in each of the racial/ethnic subgroups available in the 1988 study. In these validations, Hispanics, Asian/Pacific Islanders, and Native Americans were all initially coded as White. While this is clearly not appropriate based on theoretical grounds, it may be necessary in practice because an excited witness may be certain about Black/White distinctions, but White/Hispanic, White/Native American, and Hispanic/Native American distinctions are much more difficult to make on appearance alone.

As can be seen in Table 7, Model 2 performs best among Native Americans (coded as Whites), with a mean residual very close to 0, and an adjusted R^2 of 0.76. As might be expected, it also performs very well for the White and Black racial groups. Model 2 performs considerably less well among Hispanics and Asian/Pacific Islanders (both coded as Whites), with mean residuals of −26 and −22 mm, respectively. Stature is overestimated in these groups by an average close to 1 in. (2.54 cm).

When Hispanics are coded as Blacks, Model 2 performs much better, with a mean residual of 7.55 mm, which is very close in magnitude and direction to that for Whites and Blacks. Coding Asian/Pacific Islanders as Black also improves the performance of Model 2 in that group; however, the mean residual of 15 mm is still much larger than those for other racial groups.

From a biological point of view, the fact that coding Hispanics as Blacks substantially improves the performance of Model 2 in that group suggests that the foot size/stature relationship in Hispanics is more similar to that of Blacks than of Whites, despite the fact that Hispanics are usually pooled with other Whites in a single racial category. This is unfortunate from a forensic science point of view, since practical application of this information in a predictive model like Model 2 requires that subject descriptions clearly distinguish Hispanics from other Whites, and this distinction is unlikely to be consistently made on appearance alone. A model with individual racial indicators thus may not be very useful, even if it is statistically superior.

Table 8 reports similar validation tests for Model 7: FL FB. As expected, the adjusted R^2 for this model based on the Black/White/Hispanic subsample is slightly lower than for Model 2 (0.68 versus 0.74), and the mean residual is slightly higher (12.39 versus 6.67 mm). Model 2 also outperforms Model 7 in Native American subjects. However, the FL FB model (7) is clearly superior to Model 2 when it comes to Hispanics and Asian/Pacific Islanders, with Model 7's mean residuals approximately 79% smaller when these groups are coded as Whites, and 37 to 54% smaller when they are coded as Blacks. Surprisingly, the group most penalized by the absence of race and gender indicators in Model 7 is

Whites, where stature is underestimated by 29 mm (1.1 in.) on the average. The magnitude of the mean residual for Whites in Model 7 is comparable to those of Hispanics and Asian/Pacific Islanders in Model 2 (when they are coded as Whites), and suggests that neither model is consistently superior for all racial groups.

Both models perform well on the Whites, Blacks, and Native Americans in the validation database, although slight underestimations of stature occur in both Blacks and Whites for the model including race and gender, and moderate underestimation of stature occurs in Whites for the FL FB-only model. Validation of the race gender model on other demographic subgroups indicates that Native Americans are well accommodated by coding them as White, but that Hispanics are not. Overall, the best results are obtained when Model 2 is used, and when Hispanics and Asian/Pacific Islanders are coded as Blacks. However, this presents a problem in that descriptions based on visual data alone may not permit the distinctions between Whites and Hispanics needed to optimize Model 2.

In contrast to the race gender model, the FL FB model fits both the modeling and validation databases well with no need for witness estimates of racial/ethnic affiliation, no gross disaccommodation of any demographic subgroups, and with a relatively small loss of precision. Given this, the FL FB regression should be the model of choice for most forensic applications.

Stature Estimation from Outsole Dimensions

Model building for the estimation of stature based on boot size/width and outsole length/width is summarized in Table 9. As was the case with foot dimensions, knowledge of boot/outsole width does add significant explanatory power to the regressions, with partial F-test results as follows: boot width $F = 24.30$, df = 1818; outsole width $F = 4.53$, df = 1817. It is remarkable that these models perform nearly as well as the foot dimension models themselves. Both adjusted R^2 values of 74.7% are very close to (in fact slightly exceed) that of the FL FB model (73.5%). As can be seen in Table 10, the 95% prediction intervals for the boot/outsole models are also slightly narrower than those for the FL FB model. The fact that the boot-based estimates performed as well as the foot-based estimates in this study is encouraging. However, several factors contributing to this success do not obtain in the real forensic science world. In particular, all the boots in this study came from a single boot style and a single, extensive, sizing system. Secondly,

TABLE 9—*Model building based on boot data.*

Model	Adjusted R^2	Test	F (df)
1. Boot size and width	0.7470		
2. Boot size only	0.7398	2 vs. 1	24.30 (1,818)[a]
3. Outsole length and width	0.7474		
4. Outsole length only	0.7463	4 vs. 3	4.53 (1,817)[b]

[a] $P < 0.001$.
[b] $P < 0.05$.

TABLE 10—*Parameter estimates for recommended models based on boot data.*

Model	Length, mm	Width, mm	Intercept	95% PI Widths
2. Boot	37.37	7.68	1363.80	(±92.4, ±92.9)
4. Outsole	3.87	2.36	300.47	(±92.4, ±93.1)

the boots were professionally fit; the wearer's only contributions to size selection came as comments made to the professional on whether one size was more comfortable than another. The fact that all subjects wore the same style boot, fit by a professional, from the same extensive sizing system, works to optimize the relationship between foot dimensions and boot dimensions in this study. In the real world, shoeprints come from a variety of styles, a variety of manufacturers, and a variety of sizing systems. This introduces considerable "noise" into the estimation procedure, and so one would not necessarily expect such a tidy relationship between shoes and stature in the real world. And in fact, a recent study by Giles and Vallandigham [*2*] confirms just that.

Conclusions

The results of this study confirm earlier studies in that a strong relationship between foot length/width and stature is demonstrated. Based upon partial *F*-tests and comparisons of adjusted R^2 values and 95% prediction limits, the performance of foot-stature regressions is significantly enhanced (in a statistical sense) by the inclusion of gender and race indicators. The improvement in performance is small, however, when one considers the problems introduced by needing to know a suspect's race and gender, and when one considers the overall magnitude of 95% prediction limits in even the best equations. A simple FL FB equation is thus recommended for use with footprints.

Estimation of stature from boot size/width and outsole length/width was accomplished with virtually the same efficiency in this sample as when actual foot dimensions were used. The fact that all boots were professionally fit and came from the same style and sizing system, however, indicates that these results must be viewed as a best-case scenario, not likely to be obtained in the field.

Finally, while it is clear that there is a strong relationship between feet and stature, the results of this study firmly suggest that the models are too imprecise to be very helpful in an exclusionary situation. With best-case models having 95% prediction intervals on the order of ±86 mm (3.4 in.), the use of footprints and shoeprints in developing a suspect description is more likely to be helpful very early in a case than in an actual courtroom.

Acknowledgments

The authors are grateful to Dr. Donna Neuberg, Department of Biostatistics, Harvard School of Public Health (HSPH), for her advice and guidance in robust and exploratory statistical techniques and in linear modeling. Much of the preliminary data analysis in this project was undertaken while Dr. Gordon was doing a tutorial with her at HSPH. The authors are also grateful to Dr. Carolyn Bensel, U.S. Army Natick Research, Development, and Engineering Center, who directed the 1985 Combat Boot Fit Test. We are especially appreciative of Dr. Eugene Giles's willingness to share his own work on this topic prior to its 1991 publication. Finally, we greatly benefitted from comments made by Dr. Giles and by Mr. Thomas Greiner on drafts of this paper.

References

[*1*] Robbins, L. M., "Estimating Height and Weight from Size of Footprints," *Journal of Forensic Sciences*, Vol. 31, No. 1, Jan. 1986, pp. 143–152.

[*2*] Giles, E. and Vallandigham, P. H., "Height Estimation from Foot and Shoeprint Length," *Journal of Forensic Sciences*, Vol. 36, No. 4, July 1991, pp. 1134–1151.

[*3*] Parham, K. R., Gordon, C. C., and Bensel, C. K., "Anthropometry of the Foot and Lower Leg of U.S. Army Soldiers from Fort Jackson, South Carolina—1985," Technical Report

NATICK/TR-92/028, U.S. Army Natick Research, Development, and Engineering Center, Natick, MA, 1992.

[*4*] Gordon, C. C., Bradtmiller, B., Churchill, T., Clauser, C. E., McConville, J. T., et al., "1988 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics," Technical Report NATICK/TR-89/044, U.S. Army Natick Research, Development, and Engineering Center, Natick, MA, Sept. 1989.

[*5*] Churchill, T. D., Bradtmiller, B., and Gordon, C. C., "Computer Software Used in the U.S. Army Anthropometric Survey 1987–1988," Technical Report NATICK/TR-88/045, U.S. Army Natick Research, Development, and Engineering Center, Natick, MA, June 1988.

[*6*] Gordon, C. C. and Bradtmiller, B., "Interobserver Error in a Large Scale Anthropometric Survey," *American Journal of Human Biology*, Vol. 4, No. 2, March 1992, pp. 253–263.

[*7*] Hoaglin, D. C., Mosteller, F., and Tukey, J. W., *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York, 1983.

[*8*] Kleinbaum, D. G., Kupper, L. L., and Muller, K. E., *Applied Regression Analysis and Other Multivariable Methods*, 2nd ed., PWS-Kent Publishing Co., Boston, 1988.

[*9*] Davis, K. T., "The Foot Length to Stature Ratio: A Study of Racial Variance," master's thesis, Department of Anthropology, Texas Tech University, Lubbock, 1990.

[*10*] Neter, J. and Wasserman, W., *Applied Linear Statistical Models*, Richard D. Irwin Inc., Homewood, IL, 1974.

[*11*] Giles, E. and Klepinger, L. L., "Confidence Intervals for Estimates Based on Linear Regression in Forensic Anthropology," *Journal of Forensic Sciences*, Vol. 33, No. 5, Sept. 1988, pp. 1218–1222.

Address requests for reprints or additional information to
Dr. Claire C. Gordon
Behavioral Sciences Division
Science & Advanced Technology Directorate
U.S. Army Natick Research, Development, and Engineering Center
Natick, MA 01760-5020