Annals of Biomedical Engineering, Vol. 23, pp. 436–444, 1995
Printed in the USA. All rights reserved.

0090–6964/95 $10.50 + .00
Copyright © 1995 Biomedical Engineering Society

COMPLEX MOTOR BEHAVIORS

Two Methods for Recommending Bat Weights

A. TERRY BAHILL and MIGUEL MORNA FREITAS

Systems and Industrial Engineering, University of Arizona, Tucson, AZ

Abstract—Baseball players swung very light and very heavy bats through our instrument and the speed of the bat was recorded. These data were used to make mathematical models for each person. Then these models were coupled with equations of physics for bat-ball collisions to compute the Ideal Bat Weight for each individual. However, these calculations required the use of a sophisticated instrument that is not conveniently available to most people. So, we tried to find items in our database that correlated with Ideal Bat Weight. However, because many cells in the database were empty, we could not use traditional statistical techniques or even neural networks. Therefore, three new methods were used to estimate the missing data: (i) a neural network was trained using subjects that had no empty cells, then that neural network was used to predict the missing data, (ii) the data patching facility of a commercial software package was used, and (iii) the empty cells were filled with random numbers. Then, using these fully populated databases, several simple models were derived for recommending bat weights.

Keywords—Baseball bats, Softball bats, Ideal Bat Weight, Coefficient of restitution, Neural networks, Missing data.

INTRODUCTION

Hitting a baseball is the hardest act in all of sports (16). This act is easier if the right bat is used, but determining the best bat for any individual is difficult. Therefore, we developed a system to measure the swings of an individual, make a model for that person and recommend a specific bat weight for that person. However, this system is not conveniently available to most people. So, we used our database of the 163 people who had been measured with our system and created simple equations that can be used to recommend a bat for an individual using common parameters such as age, height, and weight.

Baseball players (*e.g.*, Babe Ruth) have used bats as heavy as 54 ounces (1.5 kg), but physicists (6,10) have said that the optimal bat weight is only 18 ounces (0.4 kg). Because no one really knew what bat weight was best, over the years there has been a lot of experimenting with bats. Most of this experimentation was illegal, because the rules say that (for professional players) the bat must be made from one solid piece of wood. To make the bat heavier, George Sisler, who was elected to the Hall Fame in 1939, pounded Victrola phonograph needles into his bat barrel, and in the 1950s Ted Kluszewski of the Cincinnati Reds hammered in tenpenny nails. To make the bat lighter, many players have drilled a hole in the end of the bat and filled it with cork. Detroit's Norm Cash admits to using a corked bat in 1961 when he won the American League batting title by hitting .361. However the corked bat may have had little to do with his success, because he presumably used a corked bat the next year when he slumped to .243. Some players have been caught publicly using doctored bats. In 1974, the bat of Graig Nettles of the Yankees shattered as it made contact, and out bounced six Super Balls. Houston's Billy Hatcher, in 1987, hit the ball, and his bat split open spraying cork all over the infield.

PHYSICS OF BAT-BALL COLLISIONS

Such experiments waste time and probably degrade performance. So, to ameliorate the bat weight conundrum, we applied principles of physics and physiology to find the best bat weight (4,5,14). First, we used the principle of conservation of momentum that states that the momentum of the bat plus the ball must be the same before and after the collision. For baseball games, the following conservation of momentum equation is appropriate:

$$W_{ball}v_{ball\text{-}before} + W_{bat}v_{bat\text{-}before} = W_{ball}v_{ball\text{-}after} + W_{bat}v_{bat\text{-}after} \quad (1)$$

where W represents weight, v represents velocity, and *before* and *after* mean before and after the bat-ball collision. Because the pitch comes toward the batter and all three

Acknowledgment—This research was supported by Worth Sports Co. and Bahill Intelligent Computer Systems.

Address correspondence to A. Terry Bahill, Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721.

(Received 19Oct94, Revised 4Jan95, Revised, Accepted 4Jan95)

other velocities are away from the batter, $v_{ball\text{-}before}$ is a negative number. This model is simple because it treats the motion of the bat as a pure translation, ignores the mass of the batter's arms, and does not consider oblique collisions.

Next we used a property associated with the energy loss in a collision, a property called the coefficient of restitution (*CoR*). One popular definition of *CoR* is the ratio of the relative speed of the objects after a collision to their relative speed before the collision:

$$CoR = -\frac{v_{ball\text{-}after} - v_{bat\text{-}after}}{v_{ball\text{-}before} - v_{bat\text{-}before}} \quad (2)$$

The coefficient of restitution depends on properties of the bat and the ball, and on the collision speed (1,6). The following equations for the coefficient of restitution, derived from experimental data given to us by Jess Heald, President of Worth Sports, are used in our computer programs. For a CU31 aluminum bat and a softball:

$$CoR = 1.17\,(0.56 - 0.001 \text{ collision speed})$$

whereas for a baseball:

$$CoR = 1.17\,(0.61 - 0.001 \text{ collision speed})$$

where the collision speed is in mph. (If the collision speed is in m/sec it must be divided by 2.24.) The original data for these equations came from experiments where balls were shot out of air cannons onto flat wooden walls. The 1.17 is in these equations because subsequent experiments by several bat manufacturers showed that ball collisions with round bats have higher coefficients of restitution than those with flat wooden walls. Aluminum bats were used for the calculations of this paper. However, perhaps the recommendations for the major leaguers should have been calculated using wooden bats, because recent unpublished experimental data collected by several bat manufacturers have shown significant differences between bats of different materials: *CoR*s of 0.51, 0.56, and 0.60 for wood, aluminum, and titanium bats, respectively.

Now the conservation of momentum and coefficient of restitution equations can be combined, and we can solve for the ball's speed after its collision with the bat to get the *batted-ball speed equation:*

$$v_{ball\text{-}after} = \frac{(W_{ball} - CoR\, W_{bat})v_{ball\text{-}before} + (W_{bat} + CoR\, W_{bat})v_{bat\text{-}before}}{W_{ball} + W_{bat}} \quad (3A)$$

There are five variables on the right side of this equation. Four of them are readily available: the weight of the ball, the weight of the bat, the coefficient of restitution of a bat-ball collision, and the speed of the pitch, which can be measured with a radar gun.

To accommodate collisions that do not occur at the center of mass of the bat, the following equation can be used:

$$v_{ball\text{-}after} = \frac{\left[CoR - \frac{W_{ball}}{W_{bat}} - \frac{W_{ball}B^2}{g\,I_0}\right](-v_{ball\text{-}before}) + (1 + \text{CoR})\,(v_{bat\text{-}before} + B\,\omega_{bat\text{-}before})}{1 + \frac{W_{ball}}{W_{bat}} + \frac{W_{ball}B^2}{g\,I_0}} \quad (3B)$$

Equation 3B is based on a model more complicated than Eq. 1, a model that treats the motion of the bat as a translation and two rotations: one around the spine and the other around a point on the bat between the hands (14). In Eq. 3B, g is the gravitational acceleration constant. There are eight variables on the right side of this equation. Six of these eight are available: the weight of the ball, the weight of the bat, the coefficient of restitution of a bat-ball collision, the moment of inertia of the bat about its center of mass (I_0), the speed of the pitch, and the distance from the center of mass of the bat to the point of the bat-ball collision (B). Most bat-ball collisions occur near the *sweet spot* of the bat, which is, however, difficult to define precisely. A bat has a center of percussion, a maximum energy transfer point, a maximum batted-ball speed point, a maximum coefficient of restitution point, a node of the primary vibration mode, a joy spot, and a least sting point. These points are different but close together. We group them together and refer to this region as the sweet spot. We measured the properties of dozens of bats and found that the sweet spot was about 85% of the distance from the knob to the end of the bat. This finding is in accord with Refs. 1, 6, 13, and 16, Worth Sports Co. (personnel communication), and Easton Aluminum Inc. (personnel communication). With this definition, the distance B is five inches for the typical wood bats used by professionals and eight inches for typical aluminum bats used by others.

MEASURING BAT SPEED

We designed and built an instrument for measuring bat speed, the Bat Chooser.[1] It has two vertical laser beams, each with associated light detectors. The forward and backward position of each subject was adjusted so that he or she swung the center of mass of the bat through the laser beams. (Beginning in 1995, the speed of the sweet spot of the bat will be measured.) The left to right position of each subject was adjusted so that bat speed was measured at the point where he or she normally makes contact with the ball, which is typically where his or her foot hits the ground. The Bat Chooser's speech synthesizer told

[1]Bat Chooser and Ideal Bat Weight are trademarks of Bahill Intelligent Computer Systems. The Bat Chooser system has been awarded U.S. patent number 5,118,102.

TABLE 1. Bats used by adults in the variable weight experiments.

Name	Description	Mass (kg)	Weight (oz)	Distance from Knob to Center of Mass (m)	Inertia with Respect to Knob (kg-m^2)
A	Regular wood bat	0.940	33.1	0.591	0.378
B	Regular wood bat	0.864	30.4	0.584	0.337
C	Wood bat filled with lead	1.217	42.8	0.622	0.538
D	Aluminum bat filled with water	1.389	48.9	0.597	0.606
E	Aluminum fungo bat	0.657	23.1	0.546	0.242
F	End loaded plastic bat	0.356	12.5	0.578	0.134

each subject, "Swing each bat as fast as you can, while still maintaining control. That is, swing as if you were trying to hit a Roger Clemens' fastball."

Typically, each subject swung six bats through the Bat Chooser. The bats ran the gamut from super light to super heavy; yet they had similar lengths and weight distributions. Most importantly, we did *not* ask the batters to swing a set of typical bats and choose the one they liked best. Rather, they swung bats unlike those that they would ever encounter in the real world and these data were used to make mathematical models for each individual. Over 100 customized bats have been built and used in our experiments over the last seven years. In many of our experiments, the six bats described in Table 1 were used. These bats were about 34 inches (0.86 m) long, with the center of mass about 23 inches (0.58 m) from the end of the handle. A different set of bats was used for little league players (5).

Each subject swung each bat through the instrument five times with the order or presentation being randomized. The bat weight and speed for each swing was recorded and used to make a mathematical model for each subject.

THE PHYSIOLOGICAL RELATIONSHIP OF FORCE AND VELOCITY

Muscle speed is inversely proportional to muscle force, as physiologists and bioengineers have shown over the last half century (2,7,8,15). When the measured bat speeds were plotted as a function of bat weight, as is shown with the lower curve in Fig. 1, a typical muscle force-velocity relationship was obtained. These data are for Leah, a member of the University of Arizona, NCAA National Champion softball team. The circles represent individual swings of the bat. There are indeed five data points for each bat; some of the circles overlap. (Bat F of Table 1 was omitted from these experiments because women softball players end their swings with the bat hitting their back and we feared that this collision might hurt the batters or the bat.)

The following three equations have been used by physiologists and bioengineers to describe the force-velocity relationship of muscle (2,7,8,15):

straight lines, $v = Ax + B$ with $A < 0$ and $B > 0$,

hyperbolas, $(x + A)(v + B) = C$ with $A > 0$, $B > 0$, and $C > 0$,

and exponentials, $v = Ae^{Bx} + C$ with $A > 0$, $B < 0$, and $C < 0$,

where v is velocity and x is force. In most of these physiological experiments the force provided by the muscle, x, was approximated with the mass of the object that was attached to the muscle. This means that the muscle mass, and the viscosity and elasticity of the load (and in our case the centripetal acceleration) was discounted (4). Amazingly these simplifications produce good models throughout a broad range of physiological experiments. Because it worked so well, we approximated the muscle force with

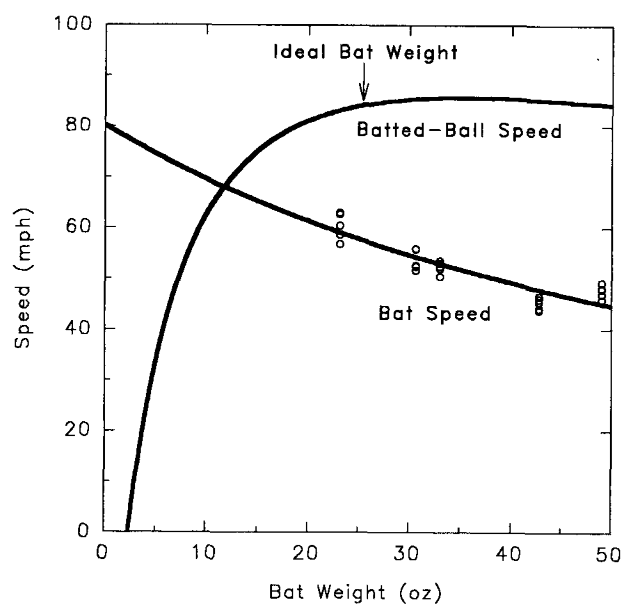

FIGURE 1. Bat speed and calculated batted-ball speed for Leah, a member of the University of Arizona softball team. One mile per hour (mph) is equal to 0.447 m/s, and 1 oz equals 28.35 g.

the weight of the bat being swung. Each of these equations has been best for some experimenters, under some conditions, with certain muscles. It was not certain that any of these equations would fit our data, because they were derived for single isolated muscles and we were trying to model whole intact human beings. We fit all three and chose the equation that gave the best fit to the data. The straight lines were fit by linear regression. The values from this fit were used to set starting points for Newton's method of steepest decent that was used to fit the exponentials. To fit the hyperbolas, we bought a commercial software package. However, we then discovered that it would only fit 12 of the 13 types of hyperbolas. Therefore, the slope and intercept of the straight line fit were used to set the starting points, and the hyperbolas were then fit by successive approximation. These fits were confirmed with sigmaplots. For our database, the hyperbolic fit was best 65% of the time, with 25% and 10% for the linear and exponential, respectively. The data of Fig. 1 were fit best with the hyperbola:

$$(W_{bat} + 70.4)(v_{bat} + 5.4) = 6{,}032 \qquad (4)$$

COUPLING PHYSICS WITH PHYSIOLOGY

Next we sought the bat weight that would make the ball leave the bat with the highest speed. This would, of course, make a potential home run go the farthest, and give a bouncing ball the maximum likelihood of getting through the infield for a hit. This bat weight is called the *maximum-batted-ball-speed bat weight*. To calculate this bat weight the muscle force-velocity relationships must be coupled with the equations of physics. The resulting equations can then be solved to find the bat weight that would allow each batter to produce the maximum batted-ball speed.

When Eq. 4, the model for Leah, is substituted into the batted-ball speed equation (for simplicity only the results for Eq. 3A will be shown), the ball speed after the collision can be plotted as a function of bat weight. This batted-ball speed curve is also shown in Fig. 1 and was calculated from the equation:

$$v_{ball\text{-}after} = \frac{(W_{ball} - CoR\ W_{bat})v_{ball\text{-}before} + (W_{bat} - CoR\ W_{bat})\left[\dfrac{6032}{W_{bat} + 70.4}\right] - 5.4}{W_{ball} + W_{bat}} \qquad (5)$$

To calculate the maximum batted-ball speed bat weight, use Eq. 5, take the derivative of $v_{ball\text{-}after}$ with respect to the bat weight, set this equal to zero, and solve for the maximum-batted-ball-speed bat weight. For this subject, the result is 31 oz, which is heavier than that used by the players on this team.

IDEAL BAT WEIGHT

However, the bat that produces the maximum batted-ball speed is not the best bat for any player. A lighter bat will give better control and more accuracy. Obviously, a trade-off must be made between batted-ball speed and controllability. Because the batted-ball speed curve (as in Fig. 1) is so flat around the point of the maximum-batted-ball-speed weight, we believe there is little advantage in using a bat as heavy as the maximum-batted-ball-speed bat weight. Therefore, we have defined the Ideal Bat Weight to be the weight at which the ball speed curve drops 1% below the maximum batted-ball speed. We believe this gives a reasonable trade-off between distance and accuracy. However, two groups of players that we measured consistently had bat speeds greater than that needed for home runs. It is not important if a home run ball lands in the fourth or the fortieth row of seats. So for these two groups, accuracy should be more important than bat speed. Therefore, for the major leaguers and the University of Arizona Softball team, we changed the parameter for the Ideal Bat Weight to be the point at which the batted-ball speed dropped by 2%. Using this criterion, the Ideal Bat Weight for the data of Fig. 1 is 25 oz. In the 1994 season, Leah actually used a 25-oz bat. She was a first team All American with a season batting average of .416.

The 1 or 2% value used to define the Ideal Bat Weight is within the range of human variability. For the professional major league players, we calculated the average and coefficient of variation for the bat speed data for the normal major league bats (the 30.4 and 33.1 oz bats of Table 1). The average of these coefficients of variation was 5.4%. This variation in bat speed data would produce a variability in the batted-ball speed of ±3.7%. This means that normal variability between consecutive swings of a normal bat would produce more than the 1 or 2% decrement used to define the Ideal Bat Weight.

The 1 or 2% value used to define the Ideal Bat Weight has no theoretical basis in physics: it is an engineering approximation that has evolved over the last 7 years. Alternative methods for determining the Ideal Bat Weight have been considered. In one, the swing accuracy *versus* bat weight is measured for individual players. Then, a performance index, such as maximizing the sum of the accuracy and batted-ball speed minus their product, is formulated. After data is collected for many players, the behavior of the performance index can be evaluated. However, all details of this alternative will be difficult to implement: so we expect to keep the 1 and 2% definitions.

In 1988, we measured the women on the University of Arizona Softball Team. For most players we recommended bats between 24 and 30 oz. In almost all cases the recommended bat was significantly lighter than the one they were using. The players replied, "But what are we to do? The lightest bat on the rack is 30 oz." We responded, "Our recommendations for tomorrow are not tied to yesterday's technology."

In 1994, we again measured the women on the University of Arizona softball team. They had won three of the last four collegiate world series and were second in the other. Again we recommended bats in the range of 24 to 31 oz, but this time most of them were already using bats within 1 oz of what we recommended.

The Ideal Bat Weight was restricted to the feasible range of 15 to 38 oz. Figure 2 shows an example at one extreme of this range. These data are for Ryan, a 9-year-old boy who is not a typical little leaguer, because his Ideal Bat Weight is very small, 15 oz. The solid circles represent the average of the five swings of each bat; the vertical bars on each circle represent the standard deviations. This plot was chosen because it shows the type of player that would profit most from switching to a light-weight bat. Players like this are often described by their coaches as being *quick* (4). His eye-hand reaction time was a relatively quick 192 ms. The data of Fig. 2 were fit best with the hyperbola:

$$(W_{bat} + 25.6)(v_{bat} + 11.6) = 2{,}170$$

For most subjects, their Ideal Bat Weight was smaller than the weight of the bat they were actually using. However, some people get higher batted-ball speeds with heavier bats. Five percent of our subjects had Ideal Bat Weights larger than 34 oz.

The Ideal Bat Weight is unique for each individual, but

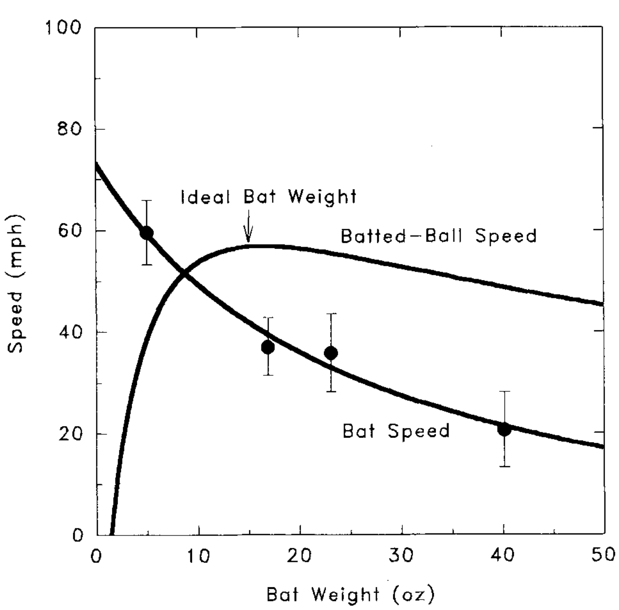

FIGURE 2. Bat speed and calculated batted-ball speed for Ryan, a 9-year-old little leaguer.

it is similar for people in the same league. Table 2 shows the means and standard deviations of Ideal Bat Weights for batters in various organized leagues. For each batter, all three force-velocity models were calculated and the one with the best fit was substituted into Eq. 3B. Depending on the team and the league, appropriate values were used for the 1 or 2% definition for the Ideal Bat Weight, the weight of the ball, the coefficient of restitution of a bat-ball collision, the moment of inertia of the bat about its center of mass (this had to be done recursively), the distance from the center of mass of the bat to the point of the bat-ball collision (B), and the speed of the pitch. These typical pitch speeds are at the high end of the range of ball speeds the bat is likely to encounter. For example, in major league baseball, the pitchers' release speeds range from 60 mph for a slow knuckleball to 100 mph for a fast fastball. But the ball loses about 10% of its speed between the pitcher's hand and the bat. So the 90 mph number is at the high end of the range of collision speeds. Al Rosen, President and CEO of the San Francisco Giants, originally suggested that we use numbers at the high end of the range.

The Ideal Bat Weight also depends on factors such as time in the season, switch hitting, *etc*. These factors are discussed in Ref. 5 and will not be discussed here.

Earlier experimenters (*e.g.*, Bahill and Karnavas [5]) concluded that the Ideal Bat Weight is specific for each individual, but it is not correlated with height, weight, age, circumference of the upper arm or any other obvious physical factors. They tried to find correlations between items in their database, but found few with even moderate correlations. The best they found was rather esoteric: the percentage superiority of the hyperbolic fit to the data over the linear fit to the data was correlated with eye-hand reaction time (correlation coefficient, $r = 0.4$) and with the slope of the straight line fit ($r = 0.5$). They considered a multiregression statistical analysis, but that approach had to be abandoned because there were a lot of missing data in the database.

A person's Ideal Bat Weight can only be determined by measuring that person's swings with an instrument like the Bat Chooser, making a model for that particular person, coupling this mathematical model with the equations of physics, and calculating the resulting Ideal Bat Weight. But most people do not have access to an instrument like the Bat Chooser, so we sought another technique to suggest the Ideal Bat Weight.

A DATABASE ANALYSIS AS AN ALTERNATIVE TO BAT CHOOSER

A database approach was used to analyze the interrelations of our data. We tried to find a good predictor of Ideal Bat Weight that does not depend on using Bat Chooser.

TABLE 2. Ideal Bat Weights.

Team	Mean Ideal Bat Weight (oz)	Standard Deviation	Typical Pitch Speed (mph)	Number of Subjects
Professional, major league	31.1	3.6	90	27
University baseball	29.4	4.2	80	9
University softball	29.7	4.3	60	19
Junior league, age 13–15	21.7	4.9	55	6
Little league, age 11–12	21.3	2.9	50	34
Little league, age 9–10	21.5	3.7	40	29
Little league, age 7–8	19.0	3.1	35	27
Slow pitch softball	25.7	3.7	20	12

A database of 36 factors on 163 subjects was constructed. Table 3 shows the name of the factor followed by the units in parentheses. Factors are grouped according to where the information came from (36 items are listed, but the four marked with asterisks were not put into the subsequent numerical database). The last item in this database, the Ideal Bat Weight, was derived from other entries in the database. First the linear, hyperbolic, or exponential Ideal Bat Weight was chosen depending on whether the straight line, hyperbola, or exponential equation fit the bat speed data best. Then, this number was restricted to the feasible range of 15 to 38 oz to produce the Ideal Bat Weight.

MANAGING MISSING VALUES

One of our early problems was that the database was not fully populated, that is, many cells were empty. There were 311 missing values, which was 6% of the total database. Three other investigators tried using neural networks to discover relationships in this database. But these missing values thwarted their efforts: neural network inputs cannot be unspecified. Assigning values for missing or uncertain neural network inputs is a difficult problem (3). The primary factors that had missing values were age, weight, extension of arms during swing, height, eye-hand reaction time, and weight of bat currently used. Some of these missing values were caused by mistakes of omission during data collection, and others were created when new questions were asked of new subjects. In the 7 years that our database has been evolving, we have gained a better understanding of the problem, and this has suggested new data that might be useful. These data are collected for the new subjects, but there is no way of obtaining them for the old subjects. This creates empty cells in the database.

Measurement of extension of arms during swing was recently introduced for use with a future model of the swing, not Eq 3. The database only had values for 17 of the 163 subjects so this column was eliminated from the database, thereby reducing the number of missing values to 3% of the total. Now 45 of the 163 subjects had data in all 31 columns.

Next, human experience was used to estimate the Age,

TABLE 3. Factors included in database approach to Ideal Bat Weight.

Questions asked of the subject:
- Name*
- Date*
- League*
- Age (years)
- Weight (lb)
- Height (in)
- Weight of bat currently used (oz)

Measurements made on the subject:
- Eye-hand reaction time (msec)
- Extension of arms during swing (cm)
- Direction of swing (1 means right handed, 0 means left handed)

Information estimated from the stated league:
- Level of play (1 means low, 10 means high)
- Typical pitch speed (mph)
- Baseball or softball* (1 means baseball, 0 means softball)

Information computed from bat swing data:
- Amount of variability in swings (ft/s)
- Amount of slowing down during tests (ft/s)

Information computed from the linear equation, $y = Ax + B$:
- Slope of the straight line (mph/oz)
- Intercept (mph)
- Mean squared error between data and line fit to data ($[ft/s]^2$)
- Maximum-batted-ball-speed bat weight (oz)
- Maximum batted-ball speed (mph)
- Linear Ideal Bat Weight (oz)

Information computed from the hyperbolic equation, $(x + A)(y + B) = C$:
- Vertical asymptote, A, (oz)
- Horizontal asymptote, B, (mph)
- Multiplicative constant, C, (oz-mph)
- Mean squared error between data and hyperbola fit to data ($[ft/s]^2$)
- Maximum-batted-ball-speed bat weight (oz)
- Maximum batted-ball speed (mph)
- Hyperbolic Ideal Bat Weight (oz)

Information computed from the exponential equation, $y = Ae^{Bx} + C$:
- Amplitude, A, (mph)
- Exponential coefficient, B, (oz^{-1})
- Offset, C, (mph)
- Mean squared error between data and exponential fit to data ($[ft/s]^2$)
- Maximum-batted-ball-speed bat weight (oz)
- Maximum batted-ball speed (mph)
- Exponential Ideal Bat Weight (oz)

Information computed from other database columns:
- Ideal Bat Weight (oz)

*Factors not included in subsequent numerical database.

and Weight of Bat Currently Used for those subjects with missing data. These estimations were made by Bahill. Since he was present during all data collection and knew many of the subjects, we felt that his estimates were better than any other type of prediction. This reduced the number of missing values to 1.4% of the total, and produced a database with 109 subjects with complete records.

Other approaches have been used to estimate missing values, for example, Montgomery (12) suggested two methods for Randomized Complete Block Design, but they are only appropriate for new experimental designs with a few missing data points, and we had an existing database with extensive missing values. Therefore, we used three new techniques to estimate the remaining missing data points.

First, the remaining 1.4% missing entries were estimated using a neural network like tool called ModelWare. The network was trained using the data of the 109 subjects who now had entries for all 31 columns. The network used 30 columns as inputs and the column with missing data points as the output. Three different networks were trained: one with weight as the output, one with height as the output, and one with eye-hand reaction time as the output. Then, these trained networks were used to predict the missing data points. This was done three times: once for weight, once for height and once for eye-hand reaction time. Thus, a 31 by 163 database was created that no longer had missing data and whose missing value predictions were based on using all the results available from Bat Chooser and all our other data.

The second technique used to estimate the remaining 1.4% empty entries was ModelWare's Patch tool. This tool fills in missing data according to ModelWare's Universal Process Modeling algorithm, which is based on a proprietary nearest neighbor approach.

The third technique generated normal random variates based on the polar method (11). For the 109 subjects with complete records, the mean and standard deviation for weight, height and eye-hand reaction time were calculated. Then, the normal random variates for each of the missing entries were generated according to the respective calculated mean and standard deviation. Thus, a third database was created where the missing values were filled with random numbers.

There were now four databases with no missing data. This allowed consideration of a multiregression statistical analysis, an approach that had been abandoned before.

NEW MODELS FOR RECOMMENDED BAT WEIGHT

These databases had 31 columns, so there were potentially 31 parameters that could be used to recommend a bat weight for each subject. However, our purpose was to develop a method that did not depend on using the Bat Chooser. Therefore, the 23 columns derived from Bat Chooser were eliminated. Then direction of swing was eliminated, because it did not help the model. Next, although it improved the model slightly, level of play was eliminated, because its evaluation was very subjective. Finally, although it helped produce a better model, the weight of bat currently used was eliminated, because the data were unreliable (43% of the data points were estimated not recorded and, for the recorded data, most of the kids first responded, "I don't know," before guessing a number), and it seemed circular to use the weight of the bat currently used to recommend the weight of bat that should be used. This left five columns of potential inputs. Using these five inputs the following "best fit" model is proposed for recommending a bat weight.

$$\text{Recommended Bat Weight} = A(\text{Age}) + B(\text{Weight}) + C(\text{Height}) + D(\text{Typical Pitch Speed}) + \text{E}(\text{Reaction Time})$$

where recommended bat weight is in ounces, age is in years (an age of 26.5 was used for all subjects over 26, because it produced a fit with the minimum error), weight is in pounds, height is in inches, typical pitch speed comes from Table 2, and reaction time is in milliseconds.

The best fit model gave a good fit to the data, but it is complex because it uses reaction time as a parameter, which restricts its usefulness because most people do not have such data. Therefore, we eliminated reaction time.

Next we sought further ways to simplify the model. The *dependence* of a parameter is given by SigmaPlot, a scientific statistical software package, and indicates the dependencies of parameters on one another: this technique suggested that weight be eliminated. *Influence coefficients* are generated by ModelWare and quantify the degree of influence exerted among system variables: this technique also suggested that weight be eliminated. Finally, using NeuralWare, a neural network was trained using these five columns as inputs to predict the Ideal Bat Weight output column. Then, one input column was removed at a time, and we noted which columns affected the error the most. This technique also suggested that weight be eliminated. Therefore, we eliminated weight. This reduction in parameters produced the "parsimonious model" (9):

$$\text{Recommended Bat Weight} = A(\text{Age}) + C(\text{Height}) + D(\text{Typical Pitch Speed})$$

The parsimonious model fits the data almost as well as the best fit model, but it is not as complex because it uses fewer parameters, and only those readily available to unsophisticated users. Finally, we again used dependencies, influence coefficients, and the neural net to eliminated age

TABLE 4. Goodness of fit of various models and databases.

	Technique Used to Estimate Missing Data			
Model	Human Knowledge (109 subjects)	Human Knowledge and Neural Net (163 subjects)	Human Knowledge and Patch Tool (163 subjects)	Human Knowledge and Random Numbers (163 subjects)
Best fit	Error: 3.34 Corr: 0.84	Error: 3.98 Corr: 0.80	Error: 4.24 Corr: 0.78	Error: 4.24 Corr: 0.78
Parsimonious	Error: 3.55 Corr: 0.83	Error: 4.11 Corr: 0.77	Error: 4.24 Corr: 0.77	Error: 4.24 Corr: 0.77
Simple	Error: 3.65 Corr: 0.82	Error: 4.37 Corr: 0.75	Error: 4.49 Corr: 0.74	Error: 4.49 Corr: 0.74

and produce the "Simple Model," which can be used to recommend a bat in a quick and easy fashion:

$$\text{Rec Bat Weight} = C(\text{Height}) + D(\text{Typical Pitch Seed})$$

The results of our statistical analysis are summarized in Table 4 in terms of the norm of the residuals, which is a measure of the root-mean-square error in fitting the data, and the coefficients of correlation between the recommended bat weight and the Ideal Bat Weight computed by Bat Chooser. These statistics were computed using SigmaPlot and were duplicated using Systat.

The smallest errors in Table 4 are associated with the 109 subject database. But use of this database was unacceptable, because the subjects had to be divided into the eight groups described in Table 2, and the 109 subject database had zero subjects in some of these groups. Of the three complete databases, the first technique for filling in missing data, training the neural network using a subset of the data to predict the missing values, yielded the best results. It has the smallest error and the highest correlation with Ideal Bat Weight values. It is interesting to note that the normal random variates technique did just as well as ModelWare's Patch tool. However, the point is not that one technique is better than the other, but rather that three different techniques were used and they all produced about the same answer. We are firm believers in using alternative procedures, because often it is not known how mathematical functions are implemented in commercial software.

When a multiple linear regression analysis of the full databases was performed, the Best fit model became:

$$\begin{aligned}\text{Recommended Bat Weight} = {} & 0.1163(\text{Age}) + 0.0696(\text{Weight}) + 0.0108(\text{Height}) \\ & + 0.1106(\text{Typical Pitch Speed}) \\ & + 0.0468(\text{Reaction Time})\end{aligned}$$

Figure 3 plots, for all 163 subjects in the database, the results from determining the recommended bat weight using the best fit model against the results using the Bat Chooser Ideal Bat Weight. This figure confirms that the best fit model is a good predictor of the Ideal Bat Weight ($r = 0.80$). This figure shows that the best fit model recommends bat weights for individuals well compared with our sophisticated instrument, the Bat Chooser. We think that these recommendations are accurate to within 2 oz, which is certainly superior to the present technique, which is, "Use the same bat as the kid down the street who hits home runs." For completeness, we now present the two simpler models derived by a multiple regression analysis on the 163 subject database.

Parsimonious model:

$$\begin{aligned}\text{Recommended Bat Weight} = {} & 0.2588(\text{Age}) + 0.2592(\text{Height}) \\ & + 0.0996(\text{Typical Pitch Speed})\end{aligned}$$

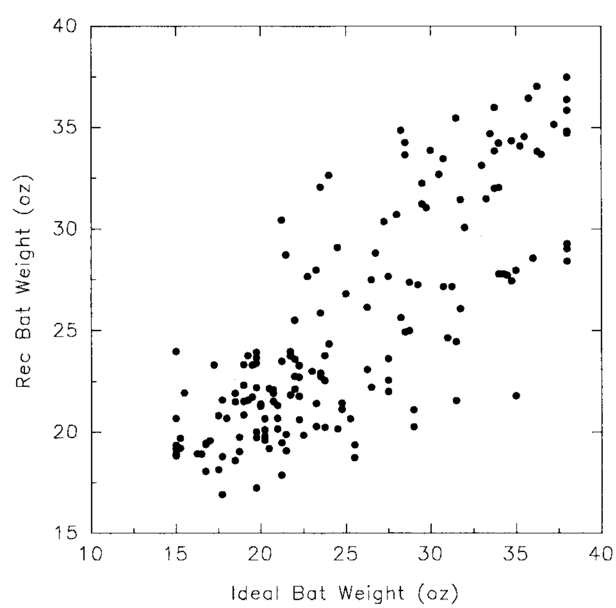

FIGURE 3. Recommended bat weight from the best fit model versus Ideal Bat Weight from the Bat Chooser.

TABLE 5. Simple integer models for recommending bat weights.

Group	Recommended Bat Weight (oz)
Baseball, major league	Height/3 + 7
Baseball, amateur	Height/3 + 6
Softball, fast pitch	Height/7 + 20
Junior league (13 & 15 years)	Height/3 + 1
Little league (11 & 12 years)	Weight/18 + 16
Little league (9 & 10 years)	Height/3 + 4
Little league (7 & 8 years)	Age*2 + 4
Softball, slow pitch	Weight/115 + 24

Age (years); height (inches); weight (pounds).

Simple Model:

$$\text{Recommended Bat Weight} = 0.3070(\text{Height}) + 0.1215(\text{Typical Pitch Speed})$$

These models should be useful: a bat can be suggested for an individual using only a simple calculator.

However, we wanted to make the models even simpler. Therefore the 163 subjects were divided into eight groups as shown in Table 2. Then we took the best fit model, restricted the number of parameters to two, restricted the parameters to be integers, and then found the values that produced the least mean-squared error between the model and the data. Thus, we created eight integer models, or rules of thumb, for use in giving quick advice to a person choosing a bat: Table 5 shows these extremely simple models. Ten years ago, no bats were available that fit these recommendations. Five years ago, there were a few. Now there are plenty of bats that fit all these recommendations.

Perhaps, the most useful entry in Table 5 shows that for a typical 9 or 10 year old:

$$\text{Recommended Bat Weight} = \frac{\text{Height}}{3} + 4$$

Clearly, this is a model that many sporting goods store employees and coaches should find very useful when asked to recommend a bat for a child.

REFERENCES

1. Adair, R. K. *The Physics of Baseball*. New York: Harper & Row, 1990, 110 pp.
2. Agarwal, G. C., and G. L. Gottlieb. Mathematical modeling and simulation of the postural control loop—part II. *CRC Crit. Rev. Biomed. Eng.* 11:113–154, 1984.
3. Bahill, A. T. *Verifying and Validating Personal Computer-Based Expert Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 113–137.
4. Bahill, A. T., and W. J. Karnavas. Determining ideal baseball bat weights using muscle force-velocity relationships. *Biol. Cybern.* 62:89–97, 1989.
5. Bahill, A. T., and W. J. Karnavas. The ideal baseball bat. *New Scientist* 130(1763):26–31, 1991.
6. Brancazio, P. *Sport Science: Physical Laws and Optimum Performance*. New York, Simon and Schuster, 1984, 400 pp.
7. Fenn, W. O., and B. S. Marsh. Muscular force at different speeds of shortening. *J. Physiol.* 85:277–297, 1935.
8. Hill, A. V. The heat of shortening and dynamic constraints of muscle. *Proc. Roy. Soc. Lond.* 126B:136–195, 1938.
9. Jefferys, W. H., and J. O. Berger. Ockham's razor and Bayesian analysis. *Am. Scientist* 80:64–72, 1992.
10. Kirkpatrick, P. Batting the ball. *Am. J. Physics* 31:606–613, 1963.
11. Law, A. M., and W. D. Kelton. *Simulation Modeling and Analysis*. New York: McGraw-Hill, 1991, 759 pp.
12. Montgomery, D. C. *Design and Analysis of Experiments*. New York: John Wiley & Sons, 1991, 649 pp.
13. Van Zandt, L. L. The dynamical theory of the baseball bat. *Am. J. Physics* 60(2):172–181, 1992.
14. Watts, R. G., and A. T. Bahill. *Keep Your Eye on the Ball: The Science and Folklore of Baseball*. New York: W. H. Freeman, 1990, 213 pp.
15. Wilke, D. R. The relation between force and velocity in human muscle. *J. Physiol.* 110:249–280, 1950.
16. Williams, T., and J. Underwood. *The Science of Hitting*. New York: Simon and Schuster, 1986, 88 pp.