A COMPREHENSIVE QUANTITATIVE AND QUALITATIVE EVALUATION OF EXTRAPOLATION OF INTRAVENOUS PHARMACOKINETIC PARAMETERS FROM RAT, DOG, AND MONKEY TO HUMANS. II. VOLUME OF DISTRIBUTION AND MEAN RESIDENCE TIME

Keith W. Ward and Brian R. Smith
Preclinical Drug Discovery, Cardiovascular & Urogenital Centre of Excellence in Drug Discovery, GlaxoSmithKline, King of Prussia, Pennsylvania

Received October 20, 2003; accepted February 23, 2004

This article is available online at http://dmd.aspetjournals.org

ABSTRACT:
Our laboratory is engaged in an ongoing analysis of a 103-compound data set containing reliable intravenous pharmacokinetic parameters in the rat, dog, monkey, and human, and we have previously reported our findings regarding extrapolation of clearance. In this article, we report our findings regarding volume of distribution and mean residence time. Various allometric and non-allometric methods were used to predict human volume of distribution based on preclinical pharmacokinetic data; clearance and volume of distribution values generated by various means were then used to estimate mean residence time. From both a quantitative and qualitative perspective, estimating human volume and mean residence time based on monkey data alone was the most accurate approach evaluated. For volume, estimation based on monkey data alone was quantitatively the least biased of all approaches evaluated. Additionally, prediction of mean residence time based on clearance and volume from the monkey was the only extrapolation method that exhibited a positive, rather than negative, bias. None of the allometric scaling approaches investigated afforded optimal predictivity for either volume or mean residence time, and neither the correlation coefficient nor the allometric exponent allowed a prospective estimation of predictive success or failure. These observations regarding volume and mean residence time confirm our earlier results with clearance, and further confirm the value of the monkey as a species for pharmacokinetic lead optimization.

The prediction of human pharmacokinetic parameters based on preclinical pharmacokinetic data has been extensively studied (Ings, 1990) and is a major goal of pharmacokineticists engaged in drug discovery. At the culmination of lead optimization efforts for a given research program, pharmacokinetic data are often available in the rat, dog, and/or monkey, and these data are then used to help predict the pharmacokinetic behavior of the lead molecule in humans. However, to date, a clear understanding of the utility of each of the major preclinical species in the quantitative or qualitative prediction of human pharmacokinetics has been limited by the lack of a comprehensive survey of primary pharmacokinetic data.

Our laboratory has recently completed an exhaustive literature survey compiling intravenous pharmacokinetic data from the rat, dog, monkey, and human for 103 nonpeptide xenobiotics. Using this data set, we are conducting a detailed analysis of interspecies pharmacokinetic relationships, and in a companion work we have discovered that the most quantitatively and qualitatively accurate prediction of human clearance was obtained by predicting human clearance as the same fraction of liver blood flow as that observed in the monkey (Ward and Smith, 2004). Using the same data set, we have also investigated interspecies relationships with respect to volume of distribution and mean residence time, and in this article we report these findings. The objectives of this work were to investigate whether any of the major species appears to be a pharmacokinetic outlier with respect to volume and mean residence time, whether data from all three species are needed for reliable extrapolation of these parameters to humans, and whether accurate prospective tools exist for evaluating probability of success of these scaling techniques.

Materials and Methods
Data Collection. Details underlying the generation of the data set used in this investigation are presented in a companion article (Ward and Smith, 2004). Briefly, an exhaustive primary literature review was conducted, and a total of 103 xenobiotics were identified with data of sufficient quality for further consideration in this extrapolation exercise. In all instances, an attempt was made to ensure that the distributional volumes used were steady-state volumes of distribution. The data set compiled was reasonably diverse with respect to human volume of distribution and mean residence time. Forty-seven compounds demonstrated a human volume of distribution less than total body water, 33 compounds had a volume between 1 and 4 times total body water, and 23 compounds had higher volumes of distribution. Additionally, 12 compounds had human mean residence times less
FIG. 1. Comparison of human volume of distribution with that of rat (A), dog (B), and monkey (C).

Volume of distribution is expressed in each species in liters per kilogram; lines represent demarcations of low (less than total body water), moderate (1 to 4 times total body water), and high (greater than 4 times total body water) volume compounds.
Fig. 2. Comparison of human volume of distribution with that of rat (A), dog (B), and monkey (C).

Volume of distribution is expressed in each species in liters per kilogram; lines represent a relationship of unity (solid line) or ± total body water (dashed lines).
than 1 h, 57 compounds demonstrated a mean residence time between 1 and 8 h, and 34 compounds had longer mean residence times.

**Data Analysis.** For the 103 xenobiotics evaluated, human volume of distribution was predicted using traditional body weight-based allometric scaling from rat, dog, and monkey or from any two of these species together to predict human volume of distribution according to the following equation (Boxenbaum, 1982): \( V = aW^b \), where \( a \) and \( b \) represent the allometric coefficient and exponent, respectively; for the three-species approach the correlation coefficient of this relationship was also calculated. Estimation of human volume of distribution from each preclinical species was also conducted by assuming that human volume would be identical to that in one of the preclinical species. Human mean residence time was predicted using the following equation (Gibaldi and Perrier, 1982): \( \frac{1}{t_{1/2}} = \frac{1}{V \cdot c_l} \), where \( V \) and \( c_l \) represent the volume of distribution and liver blood flow-based clearance, respectively; the intrinsic relationship between volume of distribution and liver blood flow-based clearance used for each molecule were those derived separately allometrically.

Before conducting any mathematical extrapolations from preclinical data to humans, the intrinsic relationship between volume of distribution in each species and in humans was first explored. Figures 1 and 2 display the relationships between volume of distribution in the rat, dog, and monkey versus human (any distributional volumes >10 times total body water were truncated in these figures to 10 times total body water for more facile data depiction). Two approaches were used to compare the preclinical species with humans: by qualitatively categorizing compounds as low (less than total body water), moderate (1 to 4 times total body water), or high (greater than 4 times total body water) volume of distribution in each species (Fig. 1); and by plotting the data in comparison with deviation from unity and evaluating outliers (compounds with volume of distribution \( \pm \) total body water of the line of unity; Fig. 2). When considering volume of distribution by category, the rat had 30 of the 103 compounds in a different volume category compared with humans (24 with a higher volume category for rat than human, and 6 with a lower category in rat than human). The dog had 28 of the 103 compounds in a different volume category compared with humans (17 with a higher volume category for dog than human, and 11 with a lower category in dog than human). Finally, the monkey had 27 of the 103 compounds in a different category compared with humans (19 with a higher volume category for monkey than human, and 8 with a lower category in monkey than human).

The interspecies differences were similar when considering predictivity based on deviation from unity (Fig. 2), with 34 compounds falling outside the \( \pm \) total body water lines for rat, 28 for dog, and 35 for monkey. Of the outliers, 76, 68, and 69\% of the rat, dog, and monkey outliers, respectively, fell above the \( \pm \) total body water line. Based on these analyses, none of the preclinical species examined appeared to be substantially more or less different from humans with respect to volume of distribution. Although not the conventional statistical mechanism for evaluating outliers, the \( \pm \) total body water analysis in Fig. 2 allows for a more intuitive outlier evaluation than a more traditional absolute fold deviation analysis.

Quantitative predictivity of human volume of distribution from the various preclinical species to humans was evaluated using standard allometric scaling from rat, dog, and monkey, rat and dog, rat and monkey, or dog and monkey to humans. Quantitative predictivity of human volume of distribution was also assessed by assuming that human volume would be identical to that in the rat, the dog, or the monkey. The accuracy of each of these methods at predicting human volume of distribution is shown in Table 1; median values are presented to ensure that the relationships are not obscured by outliers. Each of the predictive methods resulted in similar accuracy; for example, the median absolute difference between predicted and observed volume of distribution for each method ranged from 0.23 to 0.43 l/kg. However, based on each of the criteria evaluated (fold difference from observed value, variance from observed value, or absolute variance from observed value), the most quantitatively accurate prediction of human volume of distribution was achieved by assuming that human volume would be identical to that observed in the monkey. Finally, the number of times each method provided the

### TABLE 1

<table>
<thead>
<tr>
<th>Method</th>
<th>Fold Variance</th>
<th>Variance</th>
<th>Absolute Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-species allometry</td>
<td>1.03</td>
<td>0.02</td>
<td>0.27</td>
</tr>
<tr>
<td>Rat and dog allometry</td>
<td>1.15</td>
<td>0.05</td>
<td>0.31</td>
</tr>
<tr>
<td>Rat and monkey allometry</td>
<td>0.25</td>
<td>-0.27</td>
<td>0.43</td>
</tr>
<tr>
<td>Dog and monkey allometry</td>
<td>0.32</td>
<td>-0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>Rat volume</td>
<td>1.42</td>
<td>0.15</td>
<td>0.43</td>
</tr>
<tr>
<td>Dog volume</td>
<td>1.33</td>
<td>0.09</td>
<td>0.34</td>
</tr>
<tr>
<td>Monkey volume</td>
<td>1.00</td>
<td>0.00</td>
<td>0.23</td>
</tr>
<tr>
<td>Average of rat and dog</td>
<td>1.53</td>
<td>0.21</td>
<td>0.34</td>
</tr>
<tr>
<td>Average of rat and monkey</td>
<td>1.34</td>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>Average of dog and monkey</td>
<td>1.34</td>
<td>0.08</td>
<td>0.26</td>
</tr>
</tbody>
</table>

### Results

Before conducting any mathematical extrapolations from preclinical data to humans, the intrinsic relationship between volume of distribution in each species and in humans was first explored. Figures 1 and 2 display the relationships between volume of distribution in the rat, dog, and monkey versus human (any distributional volumes >10 times total body water were truncated in these figures to 10 times total body water for more facile data depiction). Two approaches were used to compare the preclinical species with humans: by qualititatively categorizing compounds as low (less than total body water), moderate (1 to 4 times total body water), or high (greater than 4 times total body water) volume of distribution in each species (Fig. 1); and by plotting the data in comparison with deviation from unity and evaluating outliers (compounds with volume of distribution \( \pm \) total body water of the line of unity; Fig. 2). When considering volume of distribution by category, the rat had 30 of the 103 compounds in a different volume category compared with humans (24 with a higher volume category for rat than human, and 6 with a lower category in rat than human). The dog had 28 of the 103 compounds in a different volume category compared with humans (17 with a higher volume category for dog than human, and 11 with a lower category in dog than human). Finally, the monkey had 27 of the 103 compounds in a different category compared with humans (19 with a higher volume category for monkey than human, and 8 with a lower category in monkey than human).

The interspecies differences were similar when considering predictivity based on deviation from unity (Fig. 2), with 34 compounds falling outside the \( \pm \) total body water lines for rat, 28 for dog, and 35 for monkey. Of the outliers, 76, 68, and 69\% of the rat, dog, and monkey outliers, respectively, fell above the \( \pm \) total body water line. Based on these analyses, none of the preclinical species examined appeared to be substantially more or less different from humans with respect to volume of distribution. Although not the conventional statistical mechanism for evaluating outliers, the \( \pm \) total body water analysis in Fig. 2 allows for a more intuitive outlier evaluation than a more traditional absolute fold deviation analysis.

Quantitative predictivity of human volume of distribution from the various preclinical species to humans was evaluated using standard allometric scaling from rat, dog, and monkey, rat and dog, rat and monkey, or dog and monkey to humans. Quantitative predictivity of human volume of distribution was also assessed by assuming that human volume would be identical to that in the rat, the dog, or the monkey. The accuracy of each of these methods at predicting human volume of distribution is shown in Table 1; median values are presented to ensure that the relationships are not obscured by outliers. Each of the predictive methods resulted in similar accuracy; for example, the median absolute difference between predicted and observed volume of distribution for each method ranged from 0.23 to 0.43 l/kg. However, based on each of the criteria evaluated (fold difference from observed value, variance from observed value, or absolute variance from observed value), the most quantitatively accurate prediction of human volume of distribution was achieved by assuming that human volume would be identical to that observed in the monkey. Finally, the number of times each method provided the

### TABLE 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Times Most Qualitatively Predictive</th>
<th>Number of Times Least Qualitatively Predictive</th>
<th>Percent Correct Qualitative Prediction</th>
<th>Percent of Incorrect Qualitative Predictions that Were Overestimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-species allometry</td>
<td>11</td>
<td>7</td>
<td>68.0</td>
<td>45.4</td>
</tr>
<tr>
<td>Rat and dog allometry</td>
<td>12</td>
<td>16</td>
<td>69.9</td>
<td>48.4</td>
</tr>
<tr>
<td>Rat and monkey allometry</td>
<td>11</td>
<td>18</td>
<td>62.1</td>
<td>10.2</td>
</tr>
<tr>
<td>Dog and monkey allometry</td>
<td>11</td>
<td>22</td>
<td>66.0</td>
<td>5.71</td>
</tr>
<tr>
<td>Rat volume</td>
<td>16</td>
<td>28</td>
<td>68.0</td>
<td>78.8</td>
</tr>
<tr>
<td>Dog volume</td>
<td>5</td>
<td>0</td>
<td>72.8</td>
<td>60.7</td>
</tr>
<tr>
<td>Monkey volume</td>
<td>19</td>
<td>11</td>
<td>73.9</td>
<td>70.4</td>
</tr>
<tr>
<td>Average of rat and dog</td>
<td>7</td>
<td>1</td>
<td>69.9</td>
<td>74.2</td>
</tr>
<tr>
<td>Average of rat and monkey</td>
<td>6</td>
<td>0</td>
<td>70.9</td>
<td>86.7</td>
</tr>
<tr>
<td>Average of dog and monkey</td>
<td>5</td>
<td>0</td>
<td>71.8</td>
<td>65.5</td>
</tr>
</tbody>
</table>
most and least accurate volume of distribution predictions was evaluated; these observations are presented in Table 2. The assumption that human volume would be identical to monkey or rat volumes provided the most accurate volume prediction in the most instances; scaling based on dog volume alone or averaging the dog and monkey volumes provided the most accurate prediction in the fewest instances. Interestingly, however, estimation of human volume as identical to that in the rat was also most frequently the least predictive method, followed by allometric scaling from dog and monkey, allometric scaling from rat and monkey, and allometric scaling from rat and dog.

As another way to represent these data, volume was next considered on a qualitative, rather than quantitative, basis. The predicted volume of distribution from each methodology and the actual human volume of distribution were classified as low, moderate, or high on the basis of the categories described above. Table 2 displays the qualitative accuracy of each of the scaling methods at predicting human volume of distribution. All the methods evaluated correctly classified volume for 62 to 73% of compounds, with human volume assumed to be the same as monkey or dog volume of distribution being the most qualitatively predictive, followed by the average of these two methods. For those instances where human volume was incorrectly classified, the different methods varied widely in whether they tended to overpredict or underpredict volume of distribution (Table 2), with allometric scaling from rat and monkey or dog and monkey tending to underpredict volume, three-species and rat and dog allometry tending to an unbiased error of prediction, and all other methods overpredicting volume. Estimating human volume of distribution based on monkey volume was the most quantitatively and qualitatively accurate method and demonstrated neither the most nor the least biased prediction.

In an attempt to further clarify the relationships between the various preclinical species, the volumes estimated from each species were compared, and then the predicted volume categories were compared (Fig. 3). When comparing rat versus dog, the two species qualitatively classified human volume the same way (although not necessarily correctly) 72 times; of the compounds where the prediction differed, the dog was correct approximately 61% of the time. Comparing rat versus monkey, a similar classification was generated 68 times; of the compounds where the prediction differed, monkey was correct in 61% of the instances. Finally, comparing dog versus monkey, the two species produced the same volume prediction 78 times, and of the compounds where the prediction differed, each species was correct approximately half the time. An additional analysis was performed to ascertain whether including or excluding either of the second nonrodent species would improve predictive quality; excluding either the dog or the monkey from the prediction made no difference in predictive quality 58% of the time, and including either of the second nonrodent species improved and worsened predictive quality to the same degree (approximately 20% of the time each; data not shown). These observations are consistent with the idea that direct estimation from either monkey or dog volume of distribution provides a qualitatively similar prediction of human volume, although monkey was somewhat more quantitatively predictive.

Finally, consideration was given to the ability to prospectively evaluate the predictive quality of the various extrapolation techniques evaluated (Fig. 4). The allometric exponent for each of the allometric methods (all species, rat and dog, rat and monkey, or dog and monkey) was calculated, as was the correlation coefficient for three-species allometry, and each was compared with qualitative predictive accuracy. For three-species allometric scaling, correlation coefficient as a measure of goodness of fit did not correspond to predictive accuracy (correlation coefficients of 0.98 and 0.97 for correct and incorrect predictions, respectively). Furthermore, the median allometric exponent did not differ substantially between compounds for which qualitative predictive success was achieved (0.824) versus those for which allometric extrapolation failed (0.737), and there was substantial overlap between the two groups, demonstrating that the allometric exponent cannot be reliably used as a prospective marker of predictive success.

With respect to mean residence time, as a function of both clearance and volume of distribution, it was anticipated that the prediction method that afforded the greatest accuracy in predicting clearance (clearance as a fraction of monkey liver blood flow) (Ward and Smith, 2004) and volume (prediction based on monkey volume of distribu-
tion; present study) would also produce the most reliable mean residence time prediction. To verify whether this was the case, human mean residence time was calculated from the various volume and clearance values predicted from the preclinical species using standard allometric scaling from rat, dog and monkey, rat and dog, rat and monkey, or dog and monkey to humans. Human mean residence time was also calculated by assuming that human volume would be identical to that of one of the preclinical species, and that clearance would be the same fraction of liver blood flow as clearance in that preclinical species. The accuracy of each of these methods at predicting human mean residence time is shown in Table 3; median values are presented to ensure that the relationships are not obscured by outliers. As anticipated, the most quantitatively accurate prediction of human mean residence time was achieved by assuming that human volume would be identical to that observed in the monkey, and that human clearance would be the same fraction of liver blood flow as that observed in the monkey. Additionally, the number of times each method provided the most and least accurate mean residence time

Fig. 4. Correlation coefficients (A) and allometric exponents (B) for compounds for which allometric scaling either correctly or incorrectly qualitatively classified human volume of distribution.
predictions was evaluated; these observations are presented in Table 4. Again, prediction, assuming that human volume would be identical to that in the monkey, and clearance, the same fraction of liver blood flow as in the monkey, resulted in the most accurate mean residence time prediction in the most instances, followed by the same methodology using rat instead of monkey; prediction based on dog and monkey or rat and monkey allometric scaling of both clearance and volume provided the most accurate prediction in the fewest instances. Prediction of mean residence time based on dog and monkey allometric scaling was by far most frequently the least predictive method, followed by rat and monkey allometric scaling and scaling from rat volume and liver blood flow.

As with volume, mean residence time was also considered on a qualitative basis. The predicted mean residence time from each methodology and the actual human mean residence time were grouped into three categories (<60 min, 60 to 480 min, and >480 min). Table 4 displays the qualitative accuracy of each of the scaling methods at predicting human mean residence time. Unlike volume of distribution, some of the methods for predicting mean residence time were substantially better than others. All of the allometric extrapolation methods demonstrated poorer qualitative predictivity than the methods based on volume and liver blood flow; dog and monkey allometry was least predictive, with only 32% correct classification, followed by rat and monkey allometry (44% correct), rat and dog allometry (61% correct), and standard three-species allometric scaling (64% correct). Overall, the most qualitatively predictive method for estimating human mean residence time was by assuming human volume would be the same as monkey volume and that human clearance would be the same fraction of liver blood flow as monkey clearance. Also, there was a strong bias with all of the allometric methodologies in the prediction of mean residence time, with the various allometric methods underestimating human mean residence time between 68 and 91% of the time. Extrapolation based on monkey data alone (the most qualitatively and quantitatively predictive method for estimating mean residence time) was neither the most nor the least biased method and underestimated human mean residence time 52% of the time.

### Discussion

Prediction of likely human pharmacokinetics of new chemical entities is an important overall target in the pharmaceutical industry, and substantial resources are expended toward this goal. In particular, at the time a molecule is selected to progress into clinical development, in vivo pharmacokinetic data are often available in the rat, dog, and/or monkey, and the pharmacokineticist is often faced with the decision of how to use these data. Also, during lead optimization, the preclinical pharmacokinetic data sometimes conflict between species, and a decision must be made regarding the most reliable species to predict human pharmacokinetics. To date, the lack of a comprehensive data set containing reliable intravenous pharmacokinetic data from the three most common preclinical species and humans has precluded an in-depth analysis of these issues. However, the present study is part of an ongoing interrogation in our laboratory of a comprehensive 103-compound data set containing intravenous pharmacokinetic parameters in the rat, dog, monkey, and human. Herein, we have arrived at some important conclusions regarding both volume of distribution and mean residence time among the preclinical species, as well as prediction of these parameters in humans.

One significant finding from this exercise is the observation that, as was observed with clearance (Ward and Smith, 2004), the monkey appears to be an important species for conducting extrapolation of volume of distribution. Furthermore, as might be anticipated, since monkey provided the most accurate prediction of both clearance and volume of distribution, predicting human mean residence time on the basis of monkey data was the most accurate of the various approaches considered here. Obviously, minimizing the use of nonhuman primates in research is an important overall objective from both a resource and an ethical perspective (Animal Procedures Committee, 2002). However, the high degree of relevance of the monkey pharmacokinetic data presented here demonstrates that, at present, the monkey cannot be replaced with either the rat or the dog in a pharmacokinetic extrapolation exercise without compromising the predictive quality of the extrapolation. Consequently, the judicious use of monkey data should continue as an important component of preclinical pharmacokinetic lead optimization and extrapolation.

Another important finding in the present investigation is the observation that generating data in two or more preclinical species does not always result in an improved prediction of volume of distribution. It has been previously suggested based on a limited data set in a variety of species that allometric scaling of volume of distribution from only two species is as predictive as that observed from three or more species (Mahmood and Balian, 1996). The data from the larger data set in the present study indicate that three-species allometry provides improved quantitative prediction compared with any two of the pre-

---

**TABLE 3**

<table>
<thead>
<tr>
<th>Method</th>
<th>Fold Variance</th>
<th>Variance min</th>
<th>Absolute Variance min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-species allometry</td>
<td>0.960</td>
<td>−4.95</td>
<td>140</td>
</tr>
<tr>
<td>Rat and dog allometry</td>
<td>0.837</td>
<td>−18.2</td>
<td>154</td>
</tr>
<tr>
<td>Rat and monkey allometry</td>
<td>0.355</td>
<td>−107</td>
<td>154</td>
</tr>
<tr>
<td>Dog and monkey allometry</td>
<td>0.078</td>
<td>−113</td>
<td>206</td>
</tr>
<tr>
<td>Rat volume and LBF</td>
<td>0.931</td>
<td>−11.3</td>
<td>172</td>
</tr>
<tr>
<td>Dog volume and LBF</td>
<td>0.797</td>
<td>−23.5</td>
<td>157</td>
</tr>
<tr>
<td>Monkey volume and LBF</td>
<td>1.09</td>
<td>11.0</td>
<td>104</td>
</tr>
</tbody>
</table>

LBF, liver blood flow.

**TABLE 4**

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Times Most Quantitatively Predictive</th>
<th>Number of Times Least Quantitatively Predictive</th>
<th>Percentage of Correct Qualitative Prediction</th>
<th>Percentage of Incorrect Qualitative Predictions that Were Overestimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-species allometry</td>
<td>13</td>
<td>7</td>
<td>64.1</td>
<td>32.4</td>
</tr>
<tr>
<td>Rat and dog allometry</td>
<td>12</td>
<td>25</td>
<td>61.1</td>
<td>25.0</td>
</tr>
<tr>
<td>Rat and monkey allometry</td>
<td>9</td>
<td>19</td>
<td>43.7</td>
<td>8.62</td>
</tr>
<tr>
<td>Dog and monkey allometry</td>
<td>4</td>
<td>46</td>
<td>32.0</td>
<td>8.57</td>
</tr>
<tr>
<td>Rat volume and LBF</td>
<td>20</td>
<td>20</td>
<td>66.0</td>
<td>48.6</td>
</tr>
<tr>
<td>Dog volume and LBF</td>
<td>16</td>
<td>0</td>
<td>69.9</td>
<td>25.8</td>
</tr>
<tr>
<td>Monkey volume and LBF</td>
<td>29</td>
<td>6</td>
<td>75.7</td>
<td>48.0</td>
</tr>
</tbody>
</table>

LBF, liver blood flow.
clinical species alone, although this was not the case from a qualitative perspective. However, the present study also shows that estimating volume based on monkey data alone provides a better prediction than either two- or three-species allometric scaling. Interestingly, however, the direction of the erroneous qualitative prediction differed substantially between methods. For all the allometric scaling-based methods, when a compound’s volume of distribution was incorrectly predicted, it was underestimated in 52 to 94% of the instances, depending on the method used. In contrast, simple volume estimation based on each preclinical species-overpredicted volume in 61 to 79% of the instances. Regardless of the method selected for extrapolation, understanding these trends should be useful in helping define the likely direction of error for any given methodology.

With respect to mean residence time, however, although extrapolation based on monkey data alone was the optimum extrapolation method, there was a clear discrimination between two- and three-species allometry. Estimation of mean residence time based on allometric scaling of clearance and volume of distribution from dog and monkey provided the quantitatively least accurate prediction for nearly half the compounds investigated, followed by allometry from rat and dog. Additionally, the two-species allometric methods provided the lowest rate of qualitative accuracy for prediction of mean residence time for all the methods evaluated. Clearly, if allometric scaling is to be used to estimate mean residence time, data from all three species should be included for the greatest likelihood of success.

The present data also highlight interesting findings with respect to traditional three-species body weight-based allometric scaling. As discussed previously (Ward and Smith, 2004), often the degree of preclinical interspecies agreement (i.e., the $r^2$ value of the linear regression) is used as a predictor of likelihood of extrapolation success. In our companion article, we have demonstrated this idea to be incorrect with respect to clearance; the present study also demonstrates that, overall, an improved mathematical correlation coefficient does not signify an improved ability of allometric scaling to correctly predict human volume (median $r^2$ value of both correct and incorrect predictions = 0.968). Additionally, the allometric exponents calculated based on the three-species scaling of the present data set contrast with the historical understanding of allometric exponents. It is generally considered that the allometric exponent for volume will revolve around 1.0 (Ings, 1990) and that extrapolations deviating from this guideline will be of lower quality, although some exceptions to this rule have been noted (Mahmood, 1999). In the present investigation, however, the average value of the allometric exponent for qualitatively correct predictions was 0.817 ± 0.280, and that for qualitatively incorrect predictions was 0.783 ± 0.399, with large overlap between the data sets. This observation indicates that allometric exponents may not be a useful prospective measure of predictive success.

The present investigation provides a reasonably comprehensive view of scaling of volume from preclinical species to humans; however, additional investigation is ongoing in our laboratory to detail further extrapolation refinements that may be of use in predicting human pharmacokinetics. For example, several investigators have established that for some xenobiotics, interspecies extrapolation of volume can be improved by the incorporation of various in vitro data such as plasma protein binding (Obach et al., 1997) or by computational approaches based on physicochemical properties of the molecule (Lombardo et al., 2002). At present, it is not possible to apply such in vitro data corrections to the compiled in vivo data in hand, due to the lack of a comprehensive and internally consistent in vitro data set. Our laboratory is currently generating these data and will report separately on their ability to improve the overall scaling of this data set, as well as the ability to predict the pharmacokinetics of these compounds based on computational methods alone.

In summary, the results from a detailed literature compilation of quality preclinical and clinical intravenous pharmacokinetic parameters demonstrate that from both a quantitative and qualitative perspective, estimating human volume and mean residence time based on monkey data alone was the most accurate approach evaluated. For volume, estimation based on monkey data alone was quantitatively the least biased of all approaches evaluated. Additionally, prediction of mean residence time based on clearance and volume from the monkey was the only extrapolation method that exhibited a positive, rather than negative, bias. None of the allometric scaling approaches investigated afforded optimal predictivity for either volume or mean residence time, and neither the correlation coefficient nor the allometric exponent allowed a prospective estimation of predictive success or failure. These observations have major implications for pharmacokinetic lead optimization and for prediction of human pharmacokinetic parameters from in vivo preclinical data, and they support the continued use of nonhuman primates in preclinical pharmacokinetics.

References