Regression analysis is the method of choice for the production of covariate-dependent reference limits. There are currently no recommendations on what sample size should be used when regression-based reference limits and confidence intervals are calculated. In this study we used Monte Carlo simulation to study a reference sample group of 374 age-dependent hemoglobin values. From this sample, 5000 random subsamples, with replacement, were constructed with 10–220 observations per sample. Regression analysis was used to estimate age-dependent 95% reference intervals for hemoglobin concentrations and erythrocyte counts. The maximum difference between mean values of the root mean square error and original values for hemoglobin was 0.05 g/L when the sample size was >60. The parameter estimators and width of reference intervals changed negligibly from the values calculated from the original sample regardless of what sample size was used. SDs and CVs for these factors changed rapidly up to a sample size of 30; after that changes were smaller. The largest and smallest absolute differences in root mean square error and width of reference interval between sample values and values calculated from the original sample were also evaluated. As expected, differences were largest in small sample sizes, and as sample size increased differences decreased. To obtain appropriate reference limits and confidence intervals, we propose the following scheme: (a) check whether the assumptions of regression analysis can be fulfilled with/without transformation of data; (b) check that the value of \( v \), which describes how the covariate value is situated in relation to both the mean value and the spread of the covariate values, does not exceed 0.1 at minimum and maximum covariate positions; and (c) if steps 1 and 2 can be accepted, the reference limits with confidence intervals can be produced by regression analysis, and the minimum acceptable sample size will be ~70.

Reference intervals are used clinically, together with additional information, as guidelines concerning the state of the patient. Both the NCCLS and IFCC have given recommendations for deriving reference values and intervals (1, 2). They have also recommended a minimum sample size of 120 for the determination of reference intervals. This number is the minimum number of values needed for construction of 90% confidence intervals by the nonparametric method. Because the reference limits are derived from a random sample they are point estimates of the true limits, which would be obtained if the whole population could be used; construction of confidence intervals gives information about the accuracy of the calculated reference limits. However, several investigators have recently produced reference limits by linear regression analysis (3–7). Linear regression analysis is particularly useful when covariate-dependency, e.g., age-dependency, exists. Provided the assumptions concerning the regression analysis are satisfied, the least-square estimators \( b_0 \) and \( b_1 \) are unbiased. According to the Gauss-Markov theorem, the least-square estimators have minimum variance among all unbiased estimators. Because they are unbiased, they tend not to overestimate or underestimate systematically. These estimators are also more precise than any of the other linear estimators (8). Continuous linear reference intervals and confidence intervals can also be constructed quite straightforwardly. Most importantly, subgrouping of data becomes unnecessary, and age-dependent reference limits can be evaluated from relatively small sample sizes. Different parametric (9–12) and nonparametric (13, 14) statistical methods have been derived to determine covariate-dependent percentiles.

According to Harris and Boyd (15), there are only a few published studies where sample size determination has been studied for reference limit estimation. Suggestions
about required sample sizes in covariate dependency have been presented only by Royston (11), who has suggested that the approximate sample size can be solved from the equation:

\[ n = \left(1 + \frac{1}{2(Z_{1-\alpha/2}^2)}\right)/(S^2/S_{del}^2) \]

where \( Z_{1-\alpha/2} \) is the quantile needed for a specified reference interval, \( S_{del} \) is the residual SD from the regression analysis, and \( S \) is the standard error of the confidence limits of the reference interval at the mean value of the covariate. Hence, if \( S \) for the 95% reference interval is 10% of \( S_{del} \) as exemplified by Royston (11), then \( n \) would be 292. This sample size, however, is unrealistically large when, for example, pediatric reference limits are calculated.

Confidence intervals are important in sample size determination because they give information about how accurate the calculated reference limits are. Covariate-based confidence intervals have been studied by Virtanen et al. (7), Royston (11), and Elveback and Taylor (16). In our previous study we verified the confidence intervals for regression-based percentiles (7). We found that 40 degrees of freedom are enough for the use of approximate confidence intervals instead of exact confidence intervals.

In the present study we determine the sample size that is needed to obtain parameter estimators and root mean square error (RMSE) to ensure that the calculated regression-based percentiles are accurate enough. To evaluate the accuracy of percentiles we used the criteria presented by Harris and Boyd [page 69; Ref. (15)], that is, the ratio of the confidence interval to the reference interval. This ratio gives information regarding whether the sample size is large enough to give narrow enough confidence intervals to ensure that the reference limits are clinically acceptable. Different sample sizes and covariate values at different positions are used to show how the ratio of the two widths varies.

**Materials and Methods**

**Subjects and Analytical Methods**

The study sample from which other samples were derived consisted of 374 children, ages 2–24 months. Some of these data (for ages 2–12 months) were also used in our earlier study (7); however, the age interval 0–2 months with a different regression model was not included in the present study. The mean age of the 374 children was 12.0 months. The minimum hemoglobin (Hb) concentration was 92 g/L, with a mean value and SD of 119.0 and 9.23 g/L, respectively. As an example of applying polynomial regression, the erythrocyte counts from the younger age group were used. These erythrocyte counts were from 99 children from newborn to 2.4 months of age. These same individuals were presented in our earlier study (7). Their mean age was 0.8 months, and their mean erythrocyte count was \( 4.42 \times 10^{12} / \text{L} \), with an SD of \( 0.82 \times 10^{12} / \text{L} \). The minimum erythrocyte count observed was \( 2.62 \times 10^{12} / \text{L} \); the maximum was \( 6.56 \times 10^{12} / \text{L} \). Overall these samples were subsamples from a larger hospital database obtained by exclusion of diagnoses that might have affected Hb or erythrocyte values (17, 18). The study protocol was officially accepted at the University Hospital of Turku and was in accordance with the Helsinki Declaration of 1975, as revised in 1983.

The Hb concentrations and erythrocyte counts were measured by Coulter Counter S-Series (S Plus VI and T-880; Coulter Electronics) or Technicon H6000 analyzers (Technicon Instruments Corp.) as described previously (18).

**Statistical Methods**

A polynomial regression model was constructed for the erythrocyte counts, and 95% reference and confidence intervals were determined. From the 374 Hb concentrations, 500 subsamples were drawn, with replacement for each of the following sample sizes: 10, 20, 30, 40, 60, 80, 100, 140, 180, and 220. A total of 5000 random subsamples were drawn. At each randomly selected age point, the mean Hb concentration was computed using the parameter estimates derived from the original sample. A random observation from a gaussian distribution with the mean value of Hb and variance (mean square error) was selected to add noise to the subsamples. For every subsample RMSE, \( b_0 \) and \( b_1 \) were calculated by the method of least squares. Mean values, SDs, and CVs were evaluated at each sample size. The mean widths of reference intervals were also determined. The RMSE and the width of the reference interval calculated from the original sample were compared with the largest and smallest values calculated from the subsamples.

The ratios of the width of the 95% confidence intervals to the width of the 95% reference intervals were calculated to evaluate the usefulness of calculated reference limits. The reasoning was the same as described by Harris and Boyd [page 69; Ref. (15)] in the univariate case:

1. The estimate of the 0.975 percentile \( (Q_0) \) calculated by the regression method in covariate value \( x_0 \) (7) is:

\[ Q_0 = b_0 + b_1 x_0 + 1.96 S_{a_{n-p}} \]

where \( b_0 \) and \( b_1 \) are parameter estimators for intercept and slope, respectively. The fractile from the gaussian distribution is 1.96, \( S \) is the RMSE, and \( a_{n-p} = ((n - p)/ (n - p - 0.5))^{1/2} \), where \( n \) is the sample size, and \( p \) is the number of parameters.

The width of the 95% reference interval equals \( 3.92 S_{a_{n-p}} \).

2. The variance of \( Q_0 \) (7) is:

\[ \sigma^2 (v + 1.96^2 (a_{n-p}^2 - 1)) \]

where \( v = (1/n + (x_0 - \bar{x})^2)/\Sigma_i (x_i - \bar{x})^2 \), and \( \bar{x} \) is the mean value of the covariate.

The width of the 95% confidence interval for reference limit is:

\[ 3.92 a_{n-p} S (v + 1.96^2 (a_{n-p}^2 - 1))^{1/2} \]
3. Hence, the ratio of the width of the confidence interval to the width of the reference interval is:

\[(v + 1.96^2(a^2_{n-p} - 1))^{1/2}\]  

(1)

As we see from Eq. 1, the ratio is dependent on the value of \(v\), and the sample size cannot be determined without knowing the value of \(v\). The value of \(v\) describes the covariate separately from other covariate values. If confidence intervals are calculated at the mean value of \(x\), then \(v = 1/n\) and the intervals are at their minimum. It is obvious that the value of Eq. 1 decreases as \(n\) increases. Different values of \(v\) and \(n\) are used to show how the ratio of the two widths varies. We chose the maximum value of \(v\) to be 0.3, even though according to Huber (19), using \(v\) values >0.2 is risky. SAS® 6.11 for Windows (SAS Institute) was used for all calculations, and Microcal Origin™ (Microcal Software) was used for graphical presentation.

**Results**

In Table 1, the mean values, SDs, and CVs of parameter estimators and RMSE are given for different sample sizes. The widths of the 95% reference intervals are also shown. The mean values of slope and intercept parameters are very close to the values calculated from the original sample regardless of sample size. From sample sizes \(\geq 60\), the maximum difference between mean values of RMSE and the original value was 0.05 g/L. SDs and CVs drop rapidly between sample sizes of 10 and 30, and after that changes are rather small (Fig. 1). The mean widths of the reference intervals are quite similar, i.e., 34–35 g/L regardless of sample size. The median differences between the largest and smallest reference intervals compared with the interval calculated from the original sample are 9.3 and 9.1 g/L, respectively (Table 2). However, when differences are observed at different sample sizes, the variation is larger. When the sample size is 10 the absolute differences between the largest and smallest reference intervals compared with the original reference interval are 29.7 and 23.5 g/L, respectively. The difference is <10 g/L when sample size is \(\geq 60\). In addition, the difference between the largest and smallest RMSE is not critical when sample size is \(>60\), i.e., the difference is ~2 g/L (Table 2).

The maximum value of \(v\) was evaluated to be 0.3. The effect of the value of \(v\) to the ratio is shown in Fig. 2 as a function of sample size. Even a sample size of 40 seems to be sufficient for calculating reference limits and confidence intervals, assuming that the maximum value of \(v\) is <0.1. For values of \(v\) of 0.3 and 0.2, the ratio is \(\geq 50\%\). Even when the sample size is near 2000 and \(v\) is 0.2, the ratio is still 45\% (not shown in Fig. 2). In Table 3, the values of the ratio when confidence intervals are calculated at mean covariate value are shown.

Both the 95% reference and confidence intervals for the study sample are shown in Fig. 3. A random sample size of 70 with 95% reference and confidence intervals is shown in Fig. 4. In this sample the largest value of \(v\) is 0.06, and the corresponding value of the ratio is \(\sim 29\%\). The 95% reference and confidence intervals for the polynomial erythrocyte model are shown in Fig. 5. In this sample the ratio of confidence interval to reference interval at the largest value of \(v\) (0.11) is \(\sim 36\%\); the ratio at the mean value of age is \(\sim 18\%\).

**Discussion**

Determinations of sample size in reference interval estimation when a covariate is present have not been widely studied. The only suggestion that we were able to find in the literature was the one by Royston (11). According to his study the approximate sample size could be derived from the equation of standard error; however, the resulting sample size is large if \(S^2/S_{\text{adj}}\) is required to be small. In this study we showed that the mean widths of reference intervals are stable and that their largest and smallest absolute differences calculated between sample values and original values are acceptable, i.e., <10 g/L when sample size is \(\geq 60\). Moreover, confidence intervals for the reference limits are not too wide at a sample size of 80, i.e.,

**Table 1. Mean values, SDs, and CVs of parameter estimators, and RMSE and width of reference interval (WRI) with different sample sizes for Hb concentrations (g/L).**

<table>
<thead>
<tr>
<th>Sample size</th>
<th>(b_0)</th>
<th>(b_1)</th>
<th>RMSE</th>
<th>WRI</th>
<th>(b_0)</th>
<th>(b_1)</th>
<th>RMSE</th>
<th>WRI</th>
<th>(b_0)</th>
<th>(b_1)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>113.5</td>
<td>0.46</td>
<td>8.50</td>
<td>34.40</td>
<td>7.15</td>
<td>0.55</td>
<td>2.14</td>
<td>8.66</td>
<td>6.30</td>
<td>120.46</td>
<td>25.18</td>
</tr>
<tr>
<td>20</td>
<td>113.8</td>
<td>0.43</td>
<td>8.79</td>
<td>34.96</td>
<td>4.44</td>
<td>0.34</td>
<td>1.48</td>
<td>5.89</td>
<td>3.90</td>
<td>80.20</td>
<td>25.18</td>
</tr>
<tr>
<td>30</td>
<td>113.4</td>
<td>0.46</td>
<td>8.81</td>
<td>34.86</td>
<td>3.67</td>
<td>0.27</td>
<td>1.16</td>
<td>4.58</td>
<td>3.23</td>
<td>59.99</td>
<td>13.15</td>
</tr>
<tr>
<td>40</td>
<td>113.5</td>
<td>0.46</td>
<td>8.70</td>
<td>34.34</td>
<td>3.41</td>
<td>0.26</td>
<td>0.97</td>
<td>3.83</td>
<td>3.00</td>
<td>55.68</td>
<td>11.16</td>
</tr>
<tr>
<td>60</td>
<td>113.6</td>
<td>0.45</td>
<td>8.78</td>
<td>34.56</td>
<td>2.56</td>
<td>0.19</td>
<td>0.85</td>
<td>3.34</td>
<td>2.25</td>
<td>42.65</td>
<td>9.66</td>
</tr>
<tr>
<td>80</td>
<td>113.8</td>
<td>0.44</td>
<td>8.78</td>
<td>34.52</td>
<td>2.12</td>
<td>0.16</td>
<td>0.70</td>
<td>2.76</td>
<td>1.86</td>
<td>36.85</td>
<td>8.00</td>
</tr>
<tr>
<td>100</td>
<td>113.7</td>
<td>0.44</td>
<td>8.81</td>
<td>34.61</td>
<td>1.86</td>
<td>0.14</td>
<td>0.61</td>
<td>2.42</td>
<td>1.63</td>
<td>31.23</td>
<td>6.98</td>
</tr>
<tr>
<td>140</td>
<td>113.8</td>
<td>0.44</td>
<td>8.82</td>
<td>34.63</td>
<td>1.69</td>
<td>0.13</td>
<td>0.53</td>
<td>2.08</td>
<td>1.48</td>
<td>28.94</td>
<td>6.00</td>
</tr>
<tr>
<td>180</td>
<td>113.7</td>
<td>0.44</td>
<td>8.81</td>
<td>34.58</td>
<td>1.37</td>
<td>0.10</td>
<td>0.48</td>
<td>1.89</td>
<td>1.21</td>
<td>23.17</td>
<td>5.47</td>
</tr>
<tr>
<td>220</td>
<td>113.5</td>
<td>0.46</td>
<td>8.80</td>
<td>34.54</td>
<td>1.33</td>
<td>0.10</td>
<td>0.41</td>
<td>1.63</td>
<td>1.17</td>
<td>21.29</td>
<td>4.71</td>
</tr>
<tr>
<td>374</td>
<td>113.6</td>
<td>0.45</td>
<td>8.83</td>
<td>34.64</td>
<td>1.33</td>
<td>0.10</td>
<td>0.41</td>
<td>1.63</td>
<td>1.17</td>
<td>21.29</td>
<td>4.71</td>
</tr>
</tbody>
</table>
the ratio of the confidence interval to the reference interval is $<20\%$ at the mean value of the covariate (Table 3).

In this study the sample size determination for the calculation of reference intervals and confidence limits was constructed for the situation of regression analysis, assuming a gaussian distribution. The method can be applied not only to data with a gaussian distribution but to data with any distribution that can be transformed to a gaussian distribution. In those situations an appropriate transformation should be used so that the assumptions of regression analysis are fulfilled, i.e., the residuals must have a gaussian distribution and their variance must be constant. Calculated reference limits and confidence intervals are retransformed to the original units. Note that when retransformation is used, confidence intervals are not of equal size around upper and lower reference limits.

Although the values in Table 1 and Fig. 1 concern Hb data, the absolute values of parameter estimators and RMSE are not important. The point of interest here is how they behave as a function of sample size. The behavior would be the same regardless of what analyte is studied under the restriction that the assumptions concerning the regression analysis are fulfilled.

When reference limits and confidence intervals are determined by regression analysis, the covariate values, such as age, should not include values that diverge considerably from other values. Extreme covariate values may cause the data to be inapplicable to regression because the value of $v$ increases and calculated confidence intervals for reference limits become so wide that they suggest that the reference intervals are not useful. Furthermore, when the value of analyte is much higher or lower than other values with similar covariate values, the

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Difference in RI</th>
<th>Difference in RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum</td>
<td>Minimum</td>
</tr>
<tr>
<td>10</td>
<td>29.7</td>
<td>23.5</td>
</tr>
<tr>
<td>20</td>
<td>22.0</td>
<td>15.8</td>
</tr>
<tr>
<td>30</td>
<td>14.5</td>
<td>11.3</td>
</tr>
<tr>
<td>40</td>
<td>13.7</td>
<td>10.2</td>
</tr>
<tr>
<td>60</td>
<td>10.3</td>
<td>11.0</td>
</tr>
<tr>
<td>80</td>
<td>8.2</td>
<td>7.9</td>
</tr>
<tr>
<td>100</td>
<td>7.0</td>
<td>6.9</td>
</tr>
<tr>
<td>140</td>
<td>7.9</td>
<td>5.9</td>
</tr>
<tr>
<td>180</td>
<td>6.1</td>
<td>6.1</td>
</tr>
<tr>
<td>220</td>
<td>4.4</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Fig. 1. CVs of slope, intercept, and RMSE and width of the reference interval.

Fig. 2. Ratio (%) of the confidence interval to the reference interval at different values of $v$ as a function of sample size.
RMSE would increase and reference intervals become wider. Huber (19) has stated that in regression analysis, data having \( v \) values >0.2 are risky to use. From Fig. 2, it can be seen that the ratio is \( \approx 50\% \) regardless of sample size when \( v \) is 0.2 or greater. Plotting and inspecting data and calculating statistical diagnostic measures [Studentized residuals, the Cook and Weisberg (20) influence statistic, high leverage points, and \( df \)betas (20)] are important to see whether the data include values that may have strong influences on the regression fit and/or RMSE and hence on the reference limit and confidence intervals.

We suggest calculating the value of \( v \) at maximum and minimum covariate values to make sure that its values are not too large. Because the parameter estimators and RMSE are estimated accurately at a rather small sample size and because the confidence intervals are not too wide at a sample size of 80, it can be concluded that a sample size of 60–80 is large enough to calculate reference limits by the regression method and to determine confidence intervals, provided that \( v \) is \( \approx 0.1 \). When \( v \) is >0.1, larger sample sizes are needed; the value of \( v \) should never exceed 0.2 because in that case the ratio of the confidence interval to the reference interval will be nearly 50\% regardless of sample size.

### References

1. National Committee for Clinical Laboratory Standards. How to define reference intervals in the clinical laboratory; approved...


