Errors of measurement for blood volume parameters: a meta-analysis
Christopher J. Gore, Will G. Hopkins and Caroline M. Burge

You might find this additional info useful...

This article cites 142 articles, 60 of which can be accessed free at:
http://jap.physiology.org/content/99/5/1745.full.html#ref-list-1

This article has been cited by 8 other HighWire hosted articles, the first 5 are:

Carbon monoxide uptake kinetics of arterial, venous and capillary blood during CO rebreathing
Laura A. Garvican, Caroline M. Burge, Amanda J. Cox, Sally A. Clark, David T. Martin and Christopher J. Gore
Exp Physiol, December 1, 2010; 95 (12): 1156-1166.
[Abstract] [Full Text] [PDF]

Alterations of systemic and muscle iron metabolism in human subjects treated with low-dose recombinant erythropoietin
Paul Robach, Stefania Recalcati, Domenico Girelli, Cecilia Gelfi, Niels J. Aachmann-Andersen, Jonas J. Thomesen, Anne M. Norgaard, Alessandra Alberghini, Natascia Campostrini, Annalisa Castagna, Agnese Viganò, Paolo Santambrogio, Tibor Kempf, Kai C. Wollert, Stéphane Moutereau, Carsten Lundby and Gaetano Cairo
Blood, June 25, 2009; 113 (26): 6707-6715.
[Abstract] [Full Text] [PDF]

Stability of hemoglobin mass over 100 days in active men
Annette Eastwood, Will G. Hopkins, Pitre C. Bourdon, Robert T. Withers and Christopher J. Gore
[Abstract] [Full Text] [PDF]

Greater uterine artery blood flow during pregnancy in multigenerational (Andean) than shorter-term (European) high-altitude residents
Megan J. Wilson, Miriam Lopez, Marco Vargas, Colleen Julian, Wilma Tellez, Armando Rodriguez, Abigail Bingham, J. Fernando Armaza, Susan Niermeyer, Mark Shriver, Enrique Vargas and Lorna G. Moore
Am J Physiol Regul Integr Comp Physiol, September 1, 2007; 293 (3): R1313-R1324.
[Abstract] [Full Text] [PDF]

Determinants of blood oxygenation during pregnancy in Andean and European residents of high altitude
Marco Vargas, Enrique Vargas, Colleen G. Julian, J. Fernando Armaza, Armando Rodriguez, Wilma Tellez, Susan Niermeyer, Megan Wilson, Esteban Parra, Mark Shriver and Lorna G. Moore
Am J Physiol Regul Integr Comp Physiol, September 1, 2007; 293 (3): R1303-R1312.
[Abstract] [Full Text] [PDF]

Updated information and services including high resolution figures, can be found at:
http://jap.physiology.org/content/99/5/1745.full.html

Additional material and information about Journal of Applied Physiology can be found at:
http://www.the-aps.org/publications/jappl

This information is current as of July 26, 2011.
Errors of measurement for blood volume parameters: a meta-analysis

Christopher J. Gore, Will G. Hopkins, and Caroline M. Burge

Errors of measurement for blood volume parameters: a meta-analysis. J Appl Physiol 99: 1745–1758, 2005. First published June 23, 2005; doi:10.1152/japplphysiol.00505.2005.—The volume of red blood cells (V_RBC) is used routinely in the diagnostic workup of polycythemia, in assessing the efficacy of erythropoietin administration, and to study factors affecting oxygen transport. However, errors of various methods of measurement of V_RBC and related parameters are not well characterized. We meta-analyzed 346 estimates of error of measurement of V_RBC for techniques based on Evans blue (V_RBC_Evans), 51chromium-labeled red blood cells (V_RBC_51Cr), and carbon monoxide (CO) rebreathing (V_RBC_CO), as well as hemoglobin mass with the carbon-monoxide method (M_Hb,CO), in athletes and active and inactive subjects undergoing various experimental and control treatments lasting minutes to months. Subject characteristics and experimental treatments had little effect on error of measurement, but measures with the smallest error showed some increase in error with increasing time between trials. Adjusted to 1 day between trials and expressed as coefficients of variation, mean errors for M_Hb,CO (2.2%; 90% confidence interval 1.4–3.5%) and V_RBC_51Cr (2.8%; 2.4–3.2%) were much less than those for V_RBC_Evans (6.7%; 4.9–9.4%) and V_RBC_CO (6.7%; 3.4–14%). Most of the error of V_RBC_Evans was due to error in measurement of volume of plasma via Evans blue dye (6.0%; 4.5–7.8%), which is the basis of V_RBC_Evans. Most of the error in V_RBC_CO was due to estimates from laboratories with a relatively large error in M_Hb,CO, the basis of V_RBC_CO. V_RBC_51Cr and M_Hb,CO are the best measures for research on blood-related changes in oxygen transport. With care, V_RBC_Evans is suitable for clinical applications of blood-volume measurement.

Address for reprint requests and other correspondence: C. J. Gore, Australian Institute of Sport, P.O. Box 176, Belconnen, ACT 2616, Australia (e-mail: chris.gore@ausport.gov.au).

http://www.jap.org
8750-7587/05 $8.00 Copyright © 2005 the American Physiological Society

1745
the intervention. Measurement errors can arise in numerous ways for the different methods used to estimate \( V_{\text{Blood}} \) or \( M_{\text{Hb}} \), as described in Table 1.

There has been no systematic review of errors in the various measures of \( V_{\text{Blood}} \) and related parameters, and it was apparent to us that the errors for \( V_{\text{RBC}} \), \( V_{\text{Plasma}} \), and \( M_{\text{Hb}} \) ranged widely, between 1 and 10%. We have, therefore, performed a meta-analysis of the errors of measurement to characterize the contribution of sampling variation, differences between laboratories, effects of subject and study characteristics, and true differences between the errors of methods used to estimate \( V_{\text{Blood}} \) values and \( M_{\text{Hb}} \).

Table 1. Comparison of sources of error for the common blood volume and hemoglobin mass methods

<table>
<thead>
<tr>
<th>Tracer</th>
<th>Evans blue</th>
<th>(^{131}I, ^{125}I)</th>
<th>(^{51}Cr, ^{99m}Tc)</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary measurement</td>
<td>Plasma dye concentration, mg/ml (plasma)</td>
<td>Counts/ml (plasma)</td>
<td>Counts/ml (whole blood)</td>
<td>HbCO %</td>
</tr>
<tr>
<td>Primary circulatory space measured by tracer</td>
<td>Albumin space</td>
<td>Albumin space</td>
<td>( V_{\text{Blood}} )</td>
<td>( M_{\text{Hb,CO}} )</td>
</tr>
<tr>
<td>Administering the tracer</td>
<td>Reliability and accuracy of tracer dose administered</td>
<td>Proportion of dose as unbound counts</td>
<td>Proportion of dose as unbound counts</td>
<td>Gas leak in re-breathing apparatus or at mouthpiece by volunteer</td>
</tr>
<tr>
<td>Potential unquantifiable loss of tracer (different on a case-by-case basis. Usually remains unknown and contributes to overall method variability)</td>
<td>Proportion of dose as unbound counts</td>
<td>Proportion of dose as unbound counts</td>
<td>Proportion of dose as unbound counts</td>
<td>Gas leak in re-breathing apparatus or at mouthpiece by volunteer</td>
</tr>
<tr>
<td>Tracer mixing kinetics</td>
<td>(\text{Counts/ml (theoretical time 0)})</td>
<td>(\Delta \text{Counts/ml (post-pre)})</td>
<td>(\Delta \text{HbCO% (post-pre)})</td>
<td>Accuracy of the CO-oximeter “span” calibration</td>
</tr>
<tr>
<td>Estimating ( V_D ) of the tracer (or equivalent)</td>
<td>(\Delta \text{Counts/ml (post-pre)})</td>
<td>(\Delta \text{HbCO% (post-pre)})</td>
<td>Reliability and precision of (HbCO) measurement (5 replicates or other criteria)</td>
<td></td>
</tr>
<tr>
<td>Steps contributing variability to the ( V_D ) estimate</td>
<td>(\Delta \text{Counts/ml (post-pre)})</td>
<td>(\Delta \text{HbCO% (post-pre)})</td>
<td>Reliability and precision of (HbCO) measurement (5 replicates or other criteria)</td>
<td></td>
</tr>
<tr>
<td>Construction of dye concentration (or counts/ml) vs. time curve. Determination of optimal number of time points and their interval</td>
<td>(\Delta \text{Counts/ml (post-pre)})</td>
<td>(\Delta \text{HbCO% (post-pre)})</td>
<td>Reliability and precision of (HbCO) measurement (5 replicates or other criteria)</td>
<td></td>
</tr>
<tr>
<td>Method selected to plot points (computer, hand) and interpolate curve (computer—spline function or other math function, eyeball)</td>
<td>(\Delta \text{Counts/ml (post-pre)})</td>
<td>(\Delta \text{HbCO% (post-pre)})</td>
<td>Reliability and precision of (HbCO) measurement (5 replicates or other criteria)</td>
<td></td>
</tr>
<tr>
<td>Mathematical model selected for back-extrapolating plasma dye conc. or count density to time zero (including all assumptions about the early rapid mixing phase), if used.</td>
<td>(\Delta \text{Counts/ml (post-pre)})</td>
<td>(\Delta \text{HbCO% (post-pre)})</td>
<td>Reliability and precision of (HbCO) measurement (5 replicates or other criteria)</td>
<td></td>
</tr>
</tbody>
</table>

| Primary method parameter | \( V_{\text{Plasma}} \) | \( V_{\text{Plasma}} \) | \( V_{\text{Blood}} \) | \( M_{\text{Hb,CO}} \) |
| Final derived parameter  | \( V_{\text{RBC}} \) | \( V_{\text{RBC}} \) | \( V_{\text{RBC}} \) | \( M_{\text{Hb,CO}} \) |
| Additional parameter required | Hct | Hct | Hct | nil |
| Primary method equation  | \( V_{\text{Plasma}} = \frac{V_D}{(dye/ml \text{dose/dye/ml plasma}) \times \text{dose volume}} \) | \( V_{\text{Plasma}} = \frac{V_D}{(\text{counts/ml dose/counts/ml plasma}) \times \text{dose volume}} \) | \( V_{\text{Blood}} = \frac{\text{counts/ml dose/counts/ml whole blood sample}}{\text{dose volume}} \times \text{dose volume} \) | \( M_{\text{Hb,CO}} = \frac{\text{CO dose}}{\text{STPD}/1.34 \text{ ml CO/g Hb}} \times \text{HbCO}\% \)
| Secondary equations to derive blood measures | \( V_{\text{Blood}} = \frac{V_{\text{Plasma}} \times (1 - \text{Hct})}{\text{Hct}} \) | \( V_{\text{RBC}} = \frac{V_{\text{Plasma}} \times (1 - \text{Hct})}{\text{Hct}} \) | \( V_{\text{RBC}} = \frac{V_{\text{RBC}}}{\text{Hct}} \) | \( V_{\text{Plasma}} = \frac{V_{\text{Blood}}}{\text{Hct}} \times \frac{M_{\text{Hb,CO}}}{\text{Hb}} \) |

\(^{131}I, ^{125}I\text{iodine; } ^{131}I, ^{125}I\text{iodine; } ^{51}Cr, ^{99m}Tc, ^{99m}Tc\text{technecium; CO, carbon monoxide; HbCO, carboxyhemoglobin; } V_{\text{Blood}} \text{, blood volume; } M_{\text{Hb,CO}} \text{, hemoglobin mass via CO; } V_{\text{Plasma}} \text{, volume of distribution; } \Delta \text{HbCO, the difference in carboxyhemoglobin levels before and after CO rebreathing; } V_{\text{Plasma, plasma volume; } V_{\text{RBC}} \text{, volume of red blood cells; Hct, hematocrit; CO dose, volume (STPD) of pure CO added to a rebreathing circuit; STPD, standard temperature and pressure dry; 1.34 is the binding capacity of hemoglobin for CO (57); } [\text{Hb}], \text{hemoglobin concentration.} \)
METHODS

Data Sources, Techniques, and Method Variations

The data used in the meta-analysis were obtained from original publications that reported values of individual subjects and from published studies whose authors provided us with their de-identified raw data (1, 2, 4–6, 10–12, 15–17, 21, 24, 25, 33, 35–37, 39–44, 47–50, 55, 63, 65, 77–80, 83, 86–91, 93, 94, 98–102, 104, 107, 108, 111, 113, 114, 117–120, 122, 132, 137–139, 141, 143, 144, 146, 148–151). Three unpublished data sets have also been included in the analysis, courtesy of the respective senior investigator responsible for data collection (Gore CJ, Slater GJ and Schmidt W, personal communications). The first two of these studies were approved by the Australian Institute of Sport Human Ethics Committee and the third was approved by the Human Ethics Committee of the Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany.

The VBlood techniques selected for the meta-analysis are those most commonly used to estimate VRBC and MHB (Table 1). We were able to obtain acceptable amounts of data for each of the following: CO rebreathing for MHB; CO rebreathing for VRBC; VRBC,51Cr; and Evans blue dye for VPlasma, VBlood, and VRBC. Some data sets included values for hematocrit (Hct) and/or hemoglobin concentration ([Hb]). Only one study (30) provided sufficient data for VPlasma, estimated using 131I-labeled albumin, and none provided data for 125I.

Modification of the CO-rebreathing technique may affect the magnitude of measurement error for MHB and its related volumes. We have previously recommended that, to minimize likely sources of error, researchers should use relatively large doses of CO, a small rebreathing volume, and at least four replicate measures of percent carboxyhemoglobin (%HbCO) (15). Others have generally used smaller doses of CO, and/or a larger rebreathing volume, and/or duplicate or single measures of %HbCO. We took this opportunity to compare the error of measurement of the “Burge and Skinner” (15) method (4–6, 15, 48–50, 122) with that of others (10, 39–41, 55, 63, 65, 117, 119, 120).

Methodological variations of the Evans blue technique could affect the magnitude of measurement error for VPlasma,Evans and its related volumes. These variations are in terms of extraction (column chromatography) or not of Evans blue dye from plasma. There are also variants that do or do not back-extrapolate a multitime point disappearance curve. Some authors do not back-extrapolate but use the simple change in plasma dye concentration obtained from one pre- and one postdye injection specimen (with post being obtained 10–20 min afterwards) (11, 24, 47, 55). The use of just one postinjection sample is justified by the statement of Greenleaf et al. (52) that a single 10-min post-dye-injection specimen gives the same VPlasma value as using the more rigorous back-extrapolation method, of which there are also several mathematical approaches (43) (Table 1). The three common method variations for VPlasma,Evans in the data that we acquired were as follows: extracted/not back-extrapolated (11, 12, 24, 44, 47, 55), not extracted/not back-extrapolated (2, 16, 21, 33, 36, 37, 42, 43, 50, 77, 78, 86–91, 93, 94, 101, 102, 104, 107, 108, 111, 118, 132, 143, 149, 150), and not extracted/back-extrapolated (17, 146).

Classically, [Hb] has been determined by pipetting blood into Drabkin’s reagent, to form cyanmethemoglobin that is measured spectrophotometrically at 540 nm (76). We distinguished between those studies that used the traditional, manual cyanmethemoglobin method to determine [Hb] and those that used fully automated analyzers.

There are no substantial variations of the CO-based method for estimation of VRBC,CO. Several altitude studies from which we obtained data (55, 98, 113, 119, 151) calculated the circulatory volumes by combining MHB with [Hb] or Hct (Table 1: secondary equations for CO); other studies (5, 40, 64, 117, 122, 127) reported only MHB,CO (15) (Table 1, primary method equation for CO).

Calculation of Error of Measurement

Studies were included in the meta-analysis if there were at least two assays of VRBC, VBlood, VPlasma, or MHB in at least three individuals receiving the same experimental treatment. The error of measurement for each blood-related measure was calculated for pairs of assays as follows (67): after log-transformation, the standard deviation of the differences was calculated, and the result was divided by \( \sqrt{2} \) and back-transformed to a CV, which was recorded along with its degrees of freedom (df = sample size – 1). These calculations were performed in a spreadsheet designed for analysis of crossovers (68). Each estimate of error was inflated by a factor \( 1 + 1/4 \) (df) to correct for small-sample bias (56).

In studies in which more than two assays were performed per subject, one assay was chosen as the reference, and the error of measurement and change in the mean were calculated for each of the other assays paired against this reference. Where two or more baseline assays were obtained before an experimental treatment, the reference was the assay closest in time to the experimental treatment. Time between pairs of assays was recorded in days.

The values of individual subjects giving rise to an estimate of error of measurement were examined for outliers in a plot of the subjects’ change score between assays against the subjects’ value of the reference assay. The following outliers were excluded from the error of measurement calculation: a subject with a 40% decline in VPlasma,CO in Poulsen et al. (111), and a subject with −40% changes in several Evans blue measures in Grover et al. (55). Individual values for the residuals and random effects in the meta-analyses (see below) were also displayed graphically for identification of other possible outliers (t values −3 or more), but none was observed. These analyses, therefore, were performed with essentially no subjective filtering of data, and any apparently large errors of measurement obtained in our analysis are not attributable to the presence of outliers.

Coding of Predictor Variables

Experimental treatments. The treatment that occurred between each pair of assays was coded either as “altitude,” “none,” or “other.” Treatments were coded as altitude for assays separated by a period of real or simulated altitude of 600–7,000 m, for exposures of 0.5–24 h/day, for durations of 0.3–40 days. Altitude treatments that included any time spent at or near sea level before, during, and/or after the altitude exposure were coded as altitude. Treatments were coded as none for pairs of assays from reliability studies, pairs of assays during a baseline period in experimental studies, and pairs of assays in any control group receiving no intervention that would be expected to change VBlood parameters. Treatments coded as other were as follows: training, ingestion of propranolol during altitude exposure, effects of menstrual cycle and/or oral contraceptive pill or placebo in combination with altitude, feeding before and after Evans blue injection, bed rest with head-down tilt and/or supine cycle ergometer training, acute exercise, living 12.8 m below the ocean surface, spaceflight, weight loss before rowing races, VPlasma expansion, heat exposure, iron supplementation, and venesection.

Subject characteristics. Sex and fitness were included in some meta-analyses. Sex was coded as a covariate representing the proportion of men in the sample (examples: 0 for all women; 0.375 for 5 women and 3 men; 1 for all men). Two studies using the 51Cr method (114, 144) and one study using the Evans blue method (143) did not state the sex of the subjects. For these studies, we assigned a sex covariate value of 0.6, which was the mean value of all of the studies. Fitness was coded as “inactive” (sedentary), “active” (but not a competitive athlete), and “athlete” (competitive). Hospital inpatients [2 studies (30, 143)] were coded as “inactive.” Four studies using the Evans blue method (24, 43, 98, 111), two using the 51Cr method (114,
144), and three using VRBC,51Cr (25, 80, 98) did not indicate physical activity or fitness of the subjects and were coded as inactive.

**Laboratories.** The effect of individual laboratories (or research groups) on error of measurement was modeled as a random effect in the meta-analysis. If a laboratory changed a particular technique substantially, a new identifier was assigned to the subsequent estimates.

**Meta-analysis**

The details of our unique, novel meta-analytic approach and its interpretation are contained in the APPENDIX. In summary, we used a mixed-model meta-analysis in which the dependent variable was the log-transformed error of measurement, the fixed effects were the method of measurement and characteristics of the study, and the random effects were within- and between-study variation. An unbiased weighting factor for each estimate was derived from the estimate’s df. Inferences about the substantiveness of true differences between two errors of measurement were made in accordance with extent of overlap of the confidence interval of their ratio with the thresholds for substantial ratios (0.9 and 1.1).

**RESULTS**

**Raw Data**

The numbers of estimates of error of measurement for each blood parameter were as follows: 69 M_Hb,CO, 32 V_Blood,Evans, 64 V_Plasma,Evans, 29 VRBC,CO, 9 VRBC,51Cr, 69 VRBC,Evans, 83 [Hb], and 57 Hct. The mean df for each parameter ranged from 5 to 10. The number of estimates for each of the methods for measurement of V_Plasma,Evans were as follows: 19 extracted/not back-extrapolated, 41 not extracted/back-extrapolated, 3 not extracted/not back-extrapolated, and 1 unclear. Thirty-six of the estimates of error for M_Hb,CO used the method of Burge and Skinner, and 33 used the other method. Of the 83 estimates for [Hb], 52 used automated methods, and 31 used the cyanmethemoglobin method. The median time between measurements giving rise to the estimates of error was 22–30 days for M_Hb,CO, V_Blood,Evans, V_Plasma,Evans, VRBC,CO, VRBC,Evans, [Hb], and Hct, but only 1 day for VRBC,51Cr. There were 159 estimates involving real or simulated altitude, 132 estimates involving other treatments, and 121 involving no treatment (from control groups, baseline pairs of measurements in treatment groups, or reliability studies). The median real or simulated altitude was 3,000 m (range 600–7,000 m), and median time spent at altitude was 12 days (0.3–40 days). The estimates were collected on 84 recreationally active, 165 athletic, and 33 used the other method. Of the 83 estimates for [Hb], 52 used automated methods, and 31 used the cyanmethemoglobin method. The median time between measurements giving rise to the estimates of error was 22–30 days for M_Hb,CO, V_Blood,Evans, V_Plasma,Evans, VRBC,CO, VRBC,Evans, [Hb], and Hct, but only 1 day for VRBC,51Cr. There were 159 estimates involving real or simulated altitude, 132 estimates involving other treatments, and 121 involving no treatment (from control groups, baseline pairs of measurements in treatment groups, or reliability studies). The median real or simulated altitude was 3,000 m (range 600–7,000 m), and median time spent at altitude was 12 days (0.3–40 days). The estimates were collected on 84 recreationally active, 165 athletic, and 161 inactive groups. For the entire data set, there was a greater proportion of men (60%) than women (40%).

**Four Comparable Hematology Parameters for the Red Cell Compartment**

The measurement errors for M_Hb,CO, V_RBC,CO, V_RBC,51Cr, and VRBC,Evans are displayed in Fig. 1 as a function of time between measurements. A 10-fold increase in time between measurements had a substantial effect on the error for M_Hb,CO (by a factor of 1.5; 90% confidence limits 1.2–1.9), trivial small effects on the errors for VRBC,51Cr (by a factor of 1.1; 1.0–1.2) and VRBC,Evans (by a factor of 1.1; 0.9–1.2), and an unclear trivial effect on the error for VRBC,CO (by a factor of 1.0; 0.6–1.6). The mean error of measurement predicted for 1 day between measurements and averaged over treatments ranged from 2.2 to 6.7% for the four methods (Table 2). M_Hb,CO had substantially less error than VRBC,51Cr (ratio of errors of M_Hb,CO to VRBC,51Cr = 0.8; 0.5–1.3), although the confidence limits indicate that the real difference was unclear. Errors for both methods were clearly less than (about one-third) those for VRBC,CO and VRBC,Evans. Estimates of the typical between-laboratory variations in the errors of measurement for M_Hb,CO, V_RBC,CO, and VRBC,Evans had considerable uncertainty, but were substantial over the range of their confidence limits: observed values (and confidence limits) were ×/×+1.8 (1.5–3.0), ×/×+1.5 (1.3–4.2), and ×/×+1.6 (1.3–2.8), respectively. Thus, for example, a new user of the M_Hb,CO method could anticipate that, in their hands, the error of measurement could easily be consistently as high as 4.0% (= 2.2 × 1.8%) or as low as 1.2% (= 2.2 ± 1.8%). The between-laboratory variation for VRBC,51Cr was not estimable, owing to a paucity of data.

The clearly smaller magnitude of measurement errors for M_Hb,CO and VRBC,51Cr implies that these blood measures are better than VRBC,CO and VRBC,Evans for quantifying effects of subject characteristics and experimental treatment on error of measurement. Only M_Hb,CO had enough estimates of error for such an analysis. The effect of a 10-fold increase in time between measurements on the error of measurement differed substantially among the three treatments (factors of 1.3, 1.1, and 1.5 for none, altitude, and other, respectively), but these differences were unclear, owing to large uncertainty in their ratios. The other effects on error for M_Hb,CO were, therefore, estimated under the assumption that the effect of a 10-fold increase in time was the same for all three treatments; this effect was reasonably clear (a factor of 1.3; confidence limits 0.9–1.8), albeit somewhat less than that provided by the simpler model above. The observed error of M_Hb,CO for women was slightly greater than that for men (ratio 1.1; 0.9–1.4), indicating that women might have substantially greater error than men but were very unlikely to have less error. Athletes were likely to have substantially greater error of measurement than active subjects (ratio 1.4; 1.0–1.9), but the greater error for athletes compared with inactive subjects was unclear, owing to a wide confidence interval for the comparison (ratio 1.5; 0.8–2.8). The active/inactive subjects comparison was also unclear (ratio 1.1; 0.6–2.0). The mean measurement error of the Burge and Skinner method of estimating M_Hb,CO (1.7%; 1.0–2.8%) was less than one-half that of the other variant (3.9%; 2.4–6.4%); the difference was definitive (ratio of Burge and Skinner/other 0.4; 0.3–0.8). There was little observed effect of treatment on error (pairwise ratios for the three treatments all 1.0), but confidence limits for the ratios (all ×/×+1.2) allowed for the possibility of small, true differences.

**V_Blood Compartments with Evans Blue**

The mean error of measurement for V_Blood,Evans, V_Plasma,Evans, and VRBC,Evans was −5–7% (Table 3), and there were no clear differences between the ratios of their errors. However, the variation in error between laboratories for these measures was substantial: ×/×+1.6 (1.4–2.0). A 10-fold increase in the time between measurements did not substantially increase the errors in V_Blood,Evans, V_Plasma,Evans, and VRBC,Evans; the ratio for the pooled data of all three volumes was ×/×+1.0 (0.9–1.1). There was little difference between errors for the pooled treatments (6.8%; 5.6–8.4%), altitude treatments (6.3%; 5.1–7.7%), and no treatment...
(6.0%; 4.9–7.3%), although confidence limits allowed for small differences between some treatment groups (range of confidence limits for ratios: 0.8–1.2).

The mean errors of measurement for \( V_{\text{Plasma,Evans}} \), determined with or without extraction and/or back-extrapolation, differed substantially but not conclusively, owing to wide confidence limits. The errors (and confidence limits) were 4.8% (2.3–10%) for not extracted/not back-extrapolated, 5.7% (3.4–9.6%) for extracted/not back-extrapolated, and 6.7% (4.5–10%) for not extracted/back-extrapolated. The estimate of the typical between-laboratory variation in the errors of measurement was substantial: \( x/\pm 1.6 \) (1.4–2.9).

![Fig. 1. Measurement error as a function of the time between measurements for 4 comparable hematological parameters: hemoglobin mass via CO rebreathing (\( M_{\text{Hb,CO}} \); ○), and volume of red blood cells via CO rebreathing (\( V_{\text{RBC,CO}} \); ×), \( ^{51} \text{Cr} \) (\( V_{\text{RBC,}^{51}\text{Cr}} \); +), and Evans blue dye (\( V_{\text{RBC,Evans}} \); △).](image)

Values of measurement error have been adjusted to remove sample-size bias. Lines of best fit are those derived from a meta-analysis, with time between measurements as the only fixed effect for each measure.

### Table 2. Measurement error of four comparable hematological parameters

<table>
<thead>
<tr>
<th>Measure</th>
<th>CV for Error, %</th>
<th>Ratio of CV Between Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( M_{\text{Hb,CO}} )</td>
</tr>
<tr>
<td>( M_{\text{Hb,CO}} )</td>
<td>2.2 (1.4–3.5)</td>
<td>1</td>
</tr>
<tr>
<td>( V_{\text{RBC,}^{51}\text{Cr}} )</td>
<td>2.8 (2.4–3.2)</td>
<td>1.2 (0.8–2.0)</td>
</tr>
<tr>
<td>( V_{\text{RBC,CO}} )</td>
<td>6.7 (3.4–14)</td>
<td>3.0 (1.3–6.9)</td>
</tr>
<tr>
<td>( V_{\text{RBC,Evans}} )</td>
<td>6.7 (4.9–9.4)</td>
<td>3.0 (1.8–5.3)</td>
</tr>
</tbody>
</table>

Data are means for both levels of treatment predicted for 1 day between measurements. Numbers in parentheses are 90% confidence limits. Errors [expressed as coefficients of variation (CV)] and their ratios were derived from a meta-analytic model in which the covariates were the parameter (4 levels: \( M_{\text{Hb,CO}} \), \( V_{\text{RBC,}^{51}\text{Cr}} \), \( V_{\text{RBC,CO}} \), \( V_{\text{RBC,Evans}} \)), its interaction with time between measurements (logarithmic metric), and its interaction with treatment between measurements (2 levels: none, other).
The substantial increase in the error for $M_{\text{HB,CO}}$ with increasing time between measurements can be accounted for partly by an increase in the contribution of biological variation. The error of measurement consists of biological variation and analytic error that are independent and combine as variances: $(\text{error of measurement})^2 = (\text{biological variation})^2 + (\text{analytic error})^2$. If we assume that the 1-day error for $M_{\text{HB,CO}}$ (2.2%) comprises minimal biological variation, the 30-day error (~4%) indicates additional error of 3.3% during this interval, irrespective of treatment. The sources of technical error (Table 1) should be independent of time between measurements, so this additional error appears to be entirely biological. An intriguing possibility is that it arises from cyclical variation in hematopoiesis with a period of approximately weeks, similar to but of lesser magnitude than that described in some hematological disorders (62). On the other hand, the additional error could be an artifact of studies of longer duration being conducted by laboratories with poorer measurement error. Frequent serial measurements of $M_{\text{HB}}$ and other blood parameters over several months should resolve this issue.

**Sources of Error**

Sources of measurement error of the common blood measurement methods are summarized in Table 1. Several of these sources of method-specific measurement error are discussed below in relation to the results obtained in the meta-analysis.

$M_{\text{HB, CO}}$. The major contributions to test-retest random analytic error for the CO rebreathing method include gas leaks in the mouthpiece, noseclip, and rebreathing system, inadequate CO dose, and using a rebreathing bag with excessive volume (15). The importance of careful attention to these sources of error is reinforced by the result that, in our hands, the Burge and Skinner method has less than one-half the error of the equivalent method used by others. An adequate dose of CO becomes particularly important if estimating %HbCO levels with commercial CO-oximeters that display readings only to a single decimal place (usually ±0.1%) (15). With progressively lower doses of CO, a 0.1% difference in the %HBCO is associated with a substantial increase in the measurement error of $M_{\text{HB,CO}}$ (15); for example, doses of 75 and 25 ml produce errors of 1.3 and 4.1%, respectively, for a woman with 600 g of hemoglobin. Investigators can perform replicate measurements with each specimen to improve the single-replicate precision (15, 18). We have successfully used a CO dose of 1.25–1.5 ml/kg in athletes with low body mass index to induce a pre- to postdose change in %HBCO of ~6.5% and performed at least five replicate measures of %HBCO (OSM-3, Radiometer, Copenhagen, Denmark) on each blood sample (50). In clinical situations, a CO dose of 1 ml/kg at sea level (aiming to increase %HBCO by ~6%) to a maximum dose of 100 ml is adequate, with appropriate dose reduction for patients with significant anemia and/or morbid obesity. Use of high-preci-
sion gas chromatography to measure %HbCO (23) allows much smaller doses of CO to be administered but requires well-trained technical staff. On the other hand, well-maintained CO-oximeters require minimal staff training and are routinely found in major hospital emergency departments and intensive care units and now increasingly in exercise physiology laboratories. The increased CO dose required when using CO-oximetry to obtain adequate precision makes no substantial difference to the safety of the method (15).

Our meta-analysis indicates that the measurement error of MHB,CO is small in most laboratories. Nevertheless, the CO method has sustained criticism of its validity, particularly the claim that inspired CO leaves the vascular space and binds to extravascular porphyrin moieties, such as myoglobin (73, 115, 126), leading to overestimation of MHB,CO (84, 100, 123, 126). However, at relatively low HbCO saturation (<15% HbCO) and at the elevated blood oxygen tension induced in the CO method (15, 48), HbCO remains stable (9), and oxygen rather than CO is taken up by myoglobin (19, 95). Observations on the effects of ischemia on deoxymyoglobin in resting muscle in the presence of 20% HbCO using magnetic resonance spectroscopy (116) support our earlier findings (15) that measures of MHB,CO do not need to be corrected for loss of CO to myoglobin. Mathematical modeling of CO uptake during 40 min of rebreathing (14) indicates that loss of CO to extravascular sites is, at most, a negligible ~1% of the CO dose over 10 min. Thus concerns of poor validity of MHB are unwarranted. Additionally, the low rate of extravascular loss of CO appears to be consistent within an individual and, therefore, will have little or no substantial effect on the random error of ~2%.

VRBC,51Cr - VRBC,51Cr is the defined criterion method of the International Committee for Standardization in Haematology (75). Our meta-analysis confirms the low measurement error of VRBC,51Cr, as concluded 25 years ago (75) and more recently (31, 32). A common misconception, however, is that the 51Cr method primarily measures VRBC. The primary estimate is actually VBlood,51Cr (32), because counts in a whole blood specimen are compared with the assay reference counts. VRBC,51Cr must then be secondarily derived from VBlood,51Cr using the Hct (Table 1). It is difficult, and potentially unethical, to perform multiple 51Cr studies on individuals simply for the sake of establishing the method measurement error. With multiple estimations in the same subject, residual radioactivity from previous estimations increases measurement error, unless compensated by progressively larger doses of 51Cr (0.5–4 MBq for one estimate, 4–20 MBq after three consecutive estimates within ~28 days). The radiation exposure is thus a concern, especially if the studies are conducted in healthy people for nondiagnostic purposes. Hence, use of the 51Cr method is rare. A variation of the chromium-labeling of red blood cells that may warrant further investigation is the use of the nonradioactive isotope 53Cr (142).

Fairbanks and colleagues (32) simulated best case scenario sources of variability affecting the measurement error of VRBC,51Cr as follows: whole blood 51Cr dose (~8,000 counts: pipetting error 1%, blood resuspension and reinjection errors 2.0%), whole blood specimen (pipetting error 1.0%, mixing error 1.0%), scintillation count error (1.1%), and Hct (1.7% biological within-person variability, 1.0% analytic variability), giving an overall estimate of the measurement error of VRBC,51Cr of 3.4%. Scintillation count errors of 1.5% (100) and 1.6% (148) have also been reported. The measurement error obtained in our meta-analysis for VRBC,51Cr was of similar magnitude to that from Fairbanks’ modeling (Table 2).

In our meta-analysis, the mean measurement error for VRBC,51Cr was substantially more than that for MHB,CO, but the confidence limits of the ratio of errors indicate that the true difference between the methods is unclear (Table 2). If the real difference is substantial, it could be due to the fact that VRBC,51Cr, unlike MHB,CO, requires Hct for its estimation (Table 1). An interesting anomaly obtained in our results, however, was that the measurement error for VRBC,51Cr is somewhat less than our estimate of the measurement error for Hct, which is not possible in reality (31). Our result can be explained as follows: the studies that we meta-analyzed for VRBC,51Cr all had measurement errors for Hct < 3%, which is near the lower 90% confidence limit for Hct. Our results are consistent with the model of Fairbanks and associates (32) that the random, biological variation in Hct from day to day is the major source of error in VRBC,51Cr (31).

VRBC,Evans. The primary estimate of the Evans blue method is VPlasma,Evans, and VRBC,Evans must be derived subsequently with the Hct (Table 1). It follows that the measurement errors for VRBC,Evans consist of those for VPlasma,Evans (6.0%) and Hct (3.2%) that combine as variances: VRBC,Evans2 = error in VPlasma,Evans2 + error in Hct2. Thus error in VRBC,Evans = (6.02 + 3.22) = 6.8%, which is consistent with our meta-analytic estimate of 6.7%. It is, therefore, clear that most of the error in VRBC,Evans derives from the measurement of VPlasma,Evans.

The existence of a wide variety of Evans blue methods is suggestive of attempts by researchers to improve method reliability. However, errors of VPlasma,Evans were not decisively different, whether a dye-extraction procedure was used (11, 12, 47, 55) or not used (43, 90, 94, 104, 108, 118), and whether (43, 88, 93, 111, 118, 143) or not (11, 12, 17, 24, 55) multiple data points were back-extrapolated to obtain the volume of distribution of Evans blue. Consequently, our analysis does not support use of one method variant over another. Analysis of multiple replicates of specimens improves analytic precision, and it is noteworthy that Gordon et al. (47), who achieved one of the lowest measurement errors for VPlasma,Evans, routinely performed triplicate spectrophotometric assays of Evans blue dye concentration. Among the other studies from which we acquired data for meta-analysis, only Branch et al. (11, 12), Chien et al. (17), and Levine et al. (personal communication) reported at least duplicate spectrophotometric assays. Different numbers of assayed replicates likely contribute to some of the variation in measurement error between laboratories, but it is unclear whether the use of multiple assays would be sufficient to make the error for VPlasma,Evans comparable to that of MHB,CO.

The Evans blue method has sustained criticism about its validity, particularly that variable loss of the dye from the vascular space causes overestimation of VPlasma as well as a relatively high measurement error (8, 29, 54). Evans blue is primarily an estimate of the albumin space (112), but the dye also binds to globulins (85), fibrinogen (8), and connective tissue (129). The albumin space includes variable flux among
the vascular, interstitial, and lymphatic spaces (8, 34); postinjection disappearance curves are consistent with a rapid mixing phase and a slow disappearance phase (143), as initially described by Gibson and Evans in 1937 (43). The rates of decline in the two phases vary between individuals (96), and substantial change within an individual between two trials would increase measurement error, especially if there were changes in extravascular flux due to increased vascular permeability. Capillary permeability to albumin increases during acute exposure to high altitude (60, 61, 92, 152) but appears to be an inconsistent response (81, 82) at the moderate altitudes used by athletes. Serum vascular endothelial growth factor [VEGF; also known as vascular permeability factor (20, 130, 131)] increases substantially at moderate altitude (3), secondary to increased oxygen-regulated gene expression (147) mediated by hypoxia inducible factor-1 (26). An increase in VEGF provides a possible means by which vascular permeability is increased at moderate altitude, and its mechanism of action appears related to disruption of endothelial tight junctions (145). The chronic increase in vascular permeability resulting from an increase in VEGF persists for at least 24 h, but not 72 h (7). Consequently, Evans blue estimates of VPlasma conducted at altitude, or 1 or 2 days after return to sea level, may be affected by an enhanced rate of loss of dye. Our analysis indicates little difference in measurement error for Evans blue volumes with various treatments, but additional error arising from a small and variable change in vascular permeability at or after altitude may have been overwhelmed by the noise of the Evans blue method.

\[ V_{RBC,CO} \]

The meta-analysis demonstrated that there was threefold greater 1-day error for \( V_{RBC,CO} \) than for MHb,CO. Part of this difference arises from the contributions that [Hb] and Hct make, along with the contribution of MHB to the error in \( V_{RBC,CO} \) (Table 1). If we assume little or no biological variation in MHB, [Hb], and Hct over 1 day, the errors in these three measures are methodological and, therefore, independent. Thus, \( (V_{RBC,CO} \text{ error})^2 = (MHB,CO \text{ error})^2 + (\text{Hct error})^2 + ([\text{Hb}] \text{ error})^2 \). In the meta-analysis, the 1-day error was 2.2% for MHB,CO, 3.2% for Hct, and 2.5% for [Hb] via the cyanmethemoglobin method. The expected error in \( V_{RBC,CO} \) is, therefore, \( \sqrt{(2.2^2 + 3.2^2 + 2.5^2)} = 4.6\% \), which is still substantially less than the meta-analyzed value of 6.7%. The simple explanation for this difference is that errors in \( V_{RBC,CO} \) come from studies with a higher than average error for MHB, which is certainly the case for the studies that provided estimates for both of these measures (55, 119, 120).

Our estimate of total error for Hct agrees well with Thirup’s (140) value of 4.2% for day-to-day variation in centrifuged micro-Hct, which he estimated as consisting of 3% biological variation and 3% analytic variation. Analytic error has been reported for automated Hct to be 2.3% and 0.8% for [Hb] (38). Using the lowest values from above, the likely minimal measurement error for \( V_{RBC,CO} \) with 1 day between measures = \( \sqrt{(2.2^2 + 2.3^2 + 0.8^2)} = 3.3\% \). This indicates that, in careful hands, the total error for \( V_{RBC,CO} \) should approach that of MHB,CO. However, the additional propagated error in \( V_{RBC,CO} \) from the measurement of Hct and [Hb] does little to support the former method if one needs to monitor small changes in the red cell compartment.

Implications for Monitoring Individuals

What are the implications of measurement error for a clinician’s uncertainty in the assessment of an individual? Applying statistical first principles, the observed value of a measurement plus or minus the error of measurement is 68% confidence limits for the true value of a normally distributed single measurement, or 52% confidence limits for a change between two such measurements; the observed value plus or minus twice the measurement error is, respectively, 95 and 84% confidence limits (66). For the clinician to be reasonably confident that an observed small but substantial positive value is not, in reality, substantially negative, it follows that the error of measurement has to be less than the least clinically important difference (66, 69, 70). In other words, to measure a signal confidently, the noise must be less than the signal (67). A measure with an error of 7% is, therefore, useful only for characterizing differences >7%. Differences in \( V_{Blood} \) of this magnitude are probably the least clinically important difference in most clinical situations, when one considers that a healthy individual can donate ~500 ml of blood from a total volume of ~5 liters with little risk. Evans blue used carefully would, therefore, be suitable in many clinical settings, as an alternative to \(^{131}\text{I} \) or \(^{125}\text{I} \) (32). Clinicians should, nevertheless, be aware that the error of measurement with Evans blue will sometimes produce unrealistically large differences in \( V_{Blood} \). Discounting such differences to an extent guided by clinical experience is sensible and an appropriate application of Bayesian reasoning (22, 70).

Implications for Controlled Trials

Adequacy of sample size is one of the most important issues in quantitative studies. With the traditional requirement of 80% chance of statistical significance at the 5% level for the smallest clinically important difference, statistical first principles show that the sample size of each group in a randomized controlled trial needs to be 32 \( d^2/e^2 \), where \( d \) is the smallest difference and \( e \) is the standard error of measurement. In studies of athletes, a 2% increase in MHB or VRBC would be important, because it would likely produce a useful change in endurance performance. A researcher interested in detecting such an increase in MHB using the CO method with an error of measurement of ~2% would, therefore, need 32 subjects in each of the control and experimental groups. Samples in the studies that we reviewed were typically less than one-half this size. Detecting the same change in \( V_{RBC} \) using an Evans blue method with an error of ~7% would require an unprecedented nine times as many subjects in each group. Less conservative approaches to sample size estimation based on adequate precision of estimation would make the usual sample sizes adequate for characterizing the smallest changes in MHB,CO, but these sample sizes would be adequate for changes in \( V_{RBC,Evans} \) only when the changes are greater than the error of measurement for \( V_{RBC,Evans} \) ~7% for the typical laboratory.

In our view, mean changes in \( V_{RBC} \) in excess of 7% are likely only following radical interventions, such as exposure to high altitude, administration of large doses of erythropoietin, and direct substantial manipulation of red cell volume by venesection or blood transfusion. There are, nevertheless, claims in the literature for even larger changes in \( V_{RBC,Evans} \) following less severe interventions, for example 4 wk of...
training at sea level (88) and 4 wk living and training at moderate altitude (16, 88). The fact that the findings were not consistent in similar studies of \( V_{RBC,51Cr} \) at higher altitude (55, 113, 151) points to type I error as an alternative explanation to such large physiological effects. Publication and other biases, such as deleting apparent outliers before analysis (e.g., a physiologically improbable large negative change in \( V_{RBC} \)), would make the rate of reporting of such large effects substantially greater than the usual type I error rate of 5%. It might help to limit such biases if researchers routinely reported the magnitude of the error of measurement in their studies, not only from any reliability study but also from the data in the control and experimental groups.

**Methods with the Smallest Error**

For researchers and clinicians seeking to improve their measurement error, we have identified the studies that demonstrated the smallest measurement errors for each of the four methods illustrated in Fig. 1.

\( M_{Hb,CO} \): The lowest measurement error of 0.9% [90% confidence limits (CL), 0.5–1.5%] was obtained by Burge and Skinner (15).

\( V_{RBC,51Cr} \): The smallest measurement error of 1.4% [90% CL = 0.6–2.9%] was obtained by Johnson et al. (79).

\( V_{RBC,Evans} \): Although the lowest estimates were ~2%, these were derived from studies of only three to four subjects [Faura and Reynefarje (33), and a subset of the 1992–2002 data from Levine and coworkers (16, 42, 86–88, 150)]. The 90% uncertainties in these estimates are, therefore, a factor of approximately \( \times/\div 2.5 \). Allowing also for regression to the mean with extreme values when there is such large sampling error, it is likely that the true error in these studies was substantially greater.

\( V_{RBC,CO} \): This method would also have acceptable error in a laboratory, with a good technique for \( M_{Hb,CO} \) and good reliability for Hct and [Hb]. The lowest errors (~2–4%) were obtained by Myhre et al. (99), although the small sample size (4–5 subjects) implies considerable uncertainty in these estimates.

Finally, it is critical that, for all measures reliant on Hct or [Hb] and to derive \( V_{RBC} \), \( V_{Plasma} \), or \( V_{Blood} \), subjects should adopt a consistent posture for at least 20 min before blood sampling (15). Venous stasis of any duration should also be avoided because it introduces substantial and unquantifiable error due to regionally increased hydrostatic pressure and localized hemococoncentration.

In conclusion, \( M_{Hb,CO} \) has error of measurement similar to that of \( V_{RBC,51Cr} \), which is considered the gold standard (32, 75, 106). Both \( V_{RBC,51Cr} \) and \( M_{Hb,CO} \) have only about one-third of the error of Evans blue dye or \( V_{RBC,CO} \) and thus are more suitable to monitor small changes in red cell volume and MHB, respectively. Given the relative ease of handling CO compared with \( 51Cr \), arising from the fact that \( M_{Hb,CO} \) is independent from biological variation in Hct, and the shorter biological half-life of CO (46), this review supports the routine use of CO rebreathing in clinical as well as research situations to monitor changes. Our results also reinforce the importance of researchers estimating and reporting the error of measurement of the method in their hands to improve the analysis and interpretation of their data.

**APPENDIX: DETAILS OF THE META-ANALYTIC APPROACH AND ITS INTERPRETATION**

**Meta-Analysis**

The main outcome from a meta-analysis is a weighted mean of values of the statistic of interest (measurement error in this instance) from the various studies, where the weighting factor is the inverse of the square of the sampling standard error of the statistic. We performed mixed-linear model meta-analyses, where effects of blood measures, treatment, and subject characteristics on the estimates were estimated as fixed effects, and the remaining unexplained true variation (heterogeneity) within and between studies was estimated as one or more random effects. Mixed-model meta-analysis is more realistic than traditional meta-analysis (in which “outlier” estimates are progressively eliminated until the test of heterogeneity is no longer statistically significant).

**Data transformation.** Meta-analysis of untransformed errors of measurement was not an option, because it would result potentially in negative values for means and confidence limits, which is not possible in reality. Analysis of the errors as variances was attractive for several reasons but suffers from the same problem. We, therefore, opted for log transformation, which was used successfully in an earlier meta-analysis of error of measurement (assessing reliability of power in physical performance tests) (72). Meta-analysis with mixed linear modeling provides confidence limits based on the assumption that random effects (including the residuals) are uniform and that any nonnormality in the individual observations is normalized in the outcome statistics by the central limit theorem. Gross departures from normality in the distribution of the individual observations are, therefore, to be avoided. We, therefore, used simulation to check that the sampling distribution of a log-transformed standard deviation derived from small sample sizes (~3) had an acceptably near-normal distribution.

**Weighting factor.** The sampling standard error for the log-transformed error of measurement was derived semi-empirically by initially assuming it was defined approximately by the 68.4% confidence limits of the standard error of measurement. These limits were derived from the usual formula involving the \( \chi^2 \) distribution and \( df \), and then converted to a single times/divide factor (\( \times/\div \) factor) by taking the square root of the upper limit divided by the lower limit. The squared log of this \( \times/\div \) factor should approximate the inverse of the weighting factor used in the meta-analysis. The accuracy of the inverse weighting factor was checked by comparing the average of its value computed for a standard deviation from each of 10,000 samples of a given size with the mean value of the variance of the log of the standard deviations of the same samples. The weighting factor was found to be biased low for small sample sizes, and a correction factor of \( 1 + 1/[2(df + 1)] \) was found by trial and error. This factor corrected the bias to within 1% for \( df \geq 2 \).

**Fixed effects.** The fixed-effects model varied with the blood measures under analysis as follows.

\( M_{Hb,Evans}, V_{RBC,Evans}, V_{RBC,CO}, \) AND \( V_{RBC,51Cr} \). For this main comparison of error of measures, a paucity of data for \( V_{RBC,51Cr} \) dictated a simple model, consisting of the interaction of measure (representing these four blood parameters) with treatment (two levels: none and other) and with \( \log_{10} \) time (the base-10 logarithm of time between pairs of measurements). Sex and fitness level were excluded from the analysis because the subtle effects of these predictors were masked by the considerably larger errors of measurement for \( V_{RBC,Evans} \) and \( V_{RBC,CO} \).

\( M_{Hb} \): A more complex fixed-effects model was used in the analysis of \( M_{Hb} \), because of its low error of measurement (see RESULTS) and a relatively large amount of available data. The predictors were sex (proportion of men in the sample), fitness (three levels: athlete, active nonathlete, and inactive), treatment (three levels: none, altitude, and other), method (Burge and Skinner and other), and \( \log_{10} \) time. The
interaction of treatment with \log_{10} time was included in a preliminary analysis.

**EVANS BLUE VOLUMES.** The predictors were measure (three levels: \(V_{\text{Blood,Evans}}\), \(V_{\text{Plasma,Evans}}\), and \(V_{\text{RBC,Evans}}\)), treatment (three levels: none, altitude, and other), and \log_{10} time.

**EVANS BLUE METHOD VARIATIONS.** The error of measurement of \(V_{\text{Plasma}}\) was the dependent variable in a model where three Evans-blue method variations (three levels: extracted/not back-extrapolated, not extracted/not back-extrapolated, and not extracted/back-extrapolated) were coded as levels of measure. The other predictors were treatment (two levels: none and other) and \log_{10} time.

([Hb] AND HCT). In separate analyses for these parameters, the models were the same as for \(M_{\text{Hb}}\), with the inclusion of method (two levels: automated and cyanmethemoglobin) for [Hb].

All meta-analyzed estimates of measurement error for a given blood parameter shown in the results are values predicted for 1 day between measurements and, where relevant, for equal contributions from each level of predictors in the model and for equal proportions of men and women. Comparisons of effects of different levels of individual predictors (e.g., no treatment vs. all other treatments) on measurement error are shown as \(X^*\) factors, derived by back-transformation of the log-transformed measurement error. For example, the difference between the error of measurement (ratio of CV between measures) for \(M_{\text{Hb,CO}}\) and that for \(V_{\text{RBC,Evans}}\) is a ratio of 3.0 (Table 2). Ratios can be reinterpreted as percent differences, for example, a ratio of 1.25 represents 25% more error, but we have opted to present all differences as ratios to minimize potential confusion with the percent units of the error of measurement. Because the time between measurements was \log_{10} transformed, its effect on measurement error is shown as a factor per 10-fold multiple of time. **Random effects.** For the main comparison of blood parameters and for the analyses of \(M_{\text{Hb,CO}}\) and Hct, the random effects coded for each measure were an identifier for the estimate (to provide the pure within-laboratory, between-estimate variation), an identifier for the laboratory technique (to provide pure between-laboratory variation), and the residual. Owing to a paucity of data for some Evans blue volumes and Evans blue methods, the between-laboratory random effect in each analysis was modeled to have the same magnitude for the three volumes and for the three method variations. The between-laboratory variance for [Hb] was negative and, therefore, set to zero; the confidence limits for the between-estimate variation were estimable only by allowing the lower confidence limit of the variance to be negative. In initial analyses, an identifier for each series of estimates of error coming from the same subjects was included as a random effect, but the paucity of data too often resulted in negative variance for one or more of the random effects. This identifier was, therefore, excluded from all analyses.

When the residual variance is scaled to unity (153), the standard deviation of the estimate identifier gives the typical variation of mean method error (CV) obtained for a method performed routinely in a specified laboratory (within-laboratory, between-estimate variation). The standard deviation of combined variances for the estimate and laboratory identifiers represents typical variation in mean method measurement error obtained between laboratories. The df for the combined variances was estimated by using the Satterthwaite (121) approximation and applied to the \(X^2\) distribution to estimate confidence limits for the combined variance (in the same manner that confidence limits for the individual variances are estimated in the statistics program). The square root of the variances and their confidence limits were back-transformed to times/divide factors. As an example, a between-laboratory typical variation of \(X^*/1.80\) obtained for a blood measure with mean measurement error (CV) of 6.7% is interpreted as follows: researchers applying the method will obtain, on average, a true CV of 6.7%, but the true value for any given researcher could be typically between 12.1% (6.7 \times 1.80%) and 3.7% (6.7 \div 1.80%). The uncertainty is actually wider, because uncertainty represented by the confidence limits for the 6.7% have not been taken into account in this example. However, since the focus of the present study is primarily examination of the fixed effects, we have shown the random effects only to illustrate the generally wide variation in the errors of measurement between assays (within laboratories) and between laboratories. These estimates of error of measurement apply only to true values or to values from large sample sizes, because sampling variation will further inflate the variation between observed estimates of error of measurement when sample sizes are small (for example, by a factor of \(X^*/1.52\) for samples of 10 subjects).

**SAS mixed-modeling procedure.** The meta-analyses were performed with the mixed-modeling procedure (Proc Mixed) in the Statistical Analysis System (version 8.2, SAS Institute, Cary, NC). Key elements of the code of the Proc Mixed step, adapted from Yang (153), are shown below for the analysis of the four related blood measures: \(M_{\text{Hb,CO}}\), \(V_{\text{RBC,Evans}}\), \(V_{\text{RBC,CO}}\), and \(V_{\text{RBC,51Cr}}\). The statements for estimating the fixed effects and for back-transforming the fixed and random effects have been omitted:

\[
\text{proc mixed};
\text{class ErrorID LabID Measure Treatment;}
\text{random ErrorID LabID /s group = 10;}\text{ddfm = sat outp = pred;}
\text{random ErrorID LabID /s group = Measure;}
\]

**Statistical inferences.** We made inferences about population (true) values of statistics via precision of estimation (using confidence limits) rather than via hypothesis testing (using \(P\) values and statistical significance). We also used 90% rather than 95% confidence limits to discourage attempts to reinterpret the limits in terms of statistical significance at the 5% level (135). Furthermore, 90% confidence limits represent adequate precision for making probabilistic inferences, because the probability that the true value is less than the lower confidence interval and more than the upper confidence interval is both only 0.05, which we interpreted as very unlikely. Inferences about the substantiveness of true differences between two errors of measurement were made by interpreting the confidence limits of the ratio in relation to the thresholds for substantial ratios (0.9 and 1.1). Ratios \(>1.1\) or \(<0.9\) were considered substantial on the basis of the impact of error on sample size: in controlled trials, sample size is inversely proportional to the square of the error of measurement (71), so an increase in error by a factor of 1.1 represents an increase in sample size of 1.21, or 21%. The difference between two errors was inferred to be clear, decisive, or conclusive in relation to one or other threshold, if the confidence interval for the ratio of the errors did not overlap the threshold. Thus one error would be clearly greater than another, if the entire confidence interval of one ratio was \(>1.1\); one error would be possibly greater than another, if the confidence interval overlapped 1.1 but not 0.9; the difference in the errors would be clearly or decisively trivial, if the confidence interval was within or abutted 0.9 to 1.1; and the difference would be unclear or indecisive, if the confidence interval overlapped 0.9 and 1.1. Overlap of a threshold that was slight relative to the width of the confidence interval was characterized by terms such as “marginal” and “unlikely,” whereas greater overlap was characterized by “possible” and “likely.”

**ACKNOWLEDGMENTS**

We thank all those who contributed de-identified raw data for us to meta-analyze. These include Clarence Allfrey, Michael Ashenden, David Branch, Kevin Davy, Birgit Friedmann, Christopher Gordon, Robert Grover, Katja Heinicke, Benjamin Levine, Jack Loeppky, Jack Reeves, Paul Robach, Philo Saunders, Walter Schmidt, Gary Slater, Ian Stewart, Jim Stray-Gundersen, and Darren Warburton.
A detailed summary table of all data sources used in the meta-analysis is available from the corresponding author on request.

REFERENCES


93. Loepky JA, Luther DK, Maes D, Riboni K, Hinghofer-Szalkay H, Charlton GA, and Icenogle MV. Plasma volume by Evans blue: effects


