The validity of serum levels as a surrogate measure of systemic exposure to metal ions in hip replacement

J. Daniel,
H. Ziaee,
P. B. Pynsent,
D. J. W. McMinn

From The McMinn Centre, Birmingham, England

Metal ions generated from joint replacements are a cause for concern. There is no consensus on the best surrogate measure of metal ion exposure. This study investigates whether serum and whole blood concentrations can be used interchangeably to report results of cobalt and chromium ion concentrations.

Concentrations of serum and whole blood were analysed in 262 concurrent specimens using high resolution inductively-coupled plasma mass-spectrometry. The agreement was assessed with normalised scatterplots, mean difference and the Bland and Altman limits of agreement.

The wide variability seen in the normalised scatterplots, in the Bland and Altman plots and the statistically significant mean differences between serum and whole blood concentrations suggest that they cannot be used interchangeably. A bias was demonstrated for both ions in the Bland-Altman plots. Regression analysis provided a possible conversion factor of 0.71 for cobalt and 0.48 for chromium. However, even when the correction factors were applied, the limits of agreement were greater than ±67% for cobalt and greater than ±85% for chromium, suggesting that serum and whole blood cannot be used interconvertibly. This suggests that serum metal concentrations are not useful as a surrogate measure of systemic metal ion exposure.

All joint replacements are subject to wear. In conventional hip replacement it has been shown that wear is a function of use[1] and that joint survival is diminished in young[2] and active[3] patients. Metal-on-metal replacements are showing good medium-term survival in young, active patients.[4-6] The main concern regarding these bearings relates to the metal ions generated from wear and their passage into the systemic circulation. The young patients who have undergone joint replacement will require a functional prosthesis for the rest of their life. The long-term effects of systemic exposure to metal ions are not yet fully understood.[7] There is a need for continued monitoring of these products.

In blood, metal ions are transported both in the plasma and within the blood cells. In the case of chromium, it has been shown that the ions within the cells are not in dynamic equilibrium with extracellular chromium and that the ratio of metal in the intra- and extracellular compartments is widely variable.[8] The concentration of metal ions in the serum correspond only to the extracellular component. Therefore, determination of whole blood concentrations are a better measure of systemic exposure to metal ions.

Monitoring has been performed[9] using whole blood, plasma, serum and erythrocytes, in addition to urine and other body fluids. The confusion surrounding the issue of metal exposure is confounded by this multiplicity of specimens used in different studies and a paucity of evidence relating to the most appropriate specimen. A panel of experts has recently conceded this inadequacy stating "to date, no study has reported a comparison of whole blood, serum, and erythrocyte levels on the same specimens in patients with metal-on-metal bearings" and that serum analysis is recommended only "due to the relative ease of analysis"[9] rather than on the basis of a scientific comparison of concurrent specimens. We have reviewed a large group of concurrent serum and whole blood specimens to determine if serum can be reliably used as a surrogate measure of metal ion exposure.

Patients and Methods

We have several ongoing cross-sectional and longitudinal studies involving metal ion analy-

■ J. Daniel, FRCS, MS(Orth), Director of Research
■ H. Ziaee, BSc(Hons), Biomedical Scientist
■ D. J. W. McMinn, FRCS, Consultant Orthopaedic Surgeon
The McMinn Centre, 25 Highfield Road, Edgbaston, Birmingham B15 3DP, UK.

■ P. B. Pynsent, PhD, Director. Research and Teaching Centre The Royal Orthopaedic Hospital, Bristol Road South, Northfield, Birmingham B31 2AP, UK.

Correspondence should be sent to Mr J. Daniel; e-mail: josephdaniel@mcminncentre.co.uk

doi:10.1302/0301-620X.89B6.18141 $2.00

J Bone Joint Surg [Br] 2007;89-B:736-41.
Received 17 May 2006; Accepted after revision 9 January 2007

Table I. Variability of metal ion levels between serum and whole blood readings of concurrent specimens

	Cobalt		**Type of replacement***			**Duration of follow-up (yrs)**				**Whole blood concentration µg/l**		
Number of specimens	**Unilateral (169)**	**Bilateral (77)**	**THR (41)**	**CORIN resurfacing (47)**	**BHR (158)**	**< 1 (53)**	**1 (85)**	**2 to 9 (74)**	**10 (34)**	**< 1 (87)**	**1 to 2 (76)**	**> 2 (83)**
Paired mean differences (µg/l)	0.6	1.3	0.8	1.4	0.7	1.9	1.2	0.4	1.7	0.5	0.6	1.4
t-test-value	< 0.0001	< 0.0001	< 0.005	< 0.005	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.01	< 0.0001	< 0.0001	< 0.0001
B-A variability (%)	*±72*	*±51.3*	*±72.2*	*±81.1*	*±59.5*	*±78.2*	*±71.6*	*±42.8*	*±68.6*	*±85.9*	*±45.4*	*±55.6*
B-A variability with correction (%)	*±76.5*	*±53.2*	*±74.3*	*±87.2*	*±62.9*	*±82.1*	*±75.8*	*±44.4*	*±74.5*	*±94.0*	*±46.8*	*±54.2*
	Chromium											
Number of specimens	**Unilateral (168)**	**Bilateral (80)**	**THR (39)**	**CORIN resurfacing (47)**	**BHR (162)**	**< 1 (62)**	**1 (80)**	**2 to 9 (72)**	**10 (34)**	**< 1 µg/l (82)**	**1 to 2 µg/l (84)**	**> 2 µg/l (82)**
Paired mean differences (µg/l)	1.6	3.5	1.4	3.0	2.0	2.5	2.5	1.5	1.2	1.0	1.4	3.9
t-test-value	< 0.0001	< 0.0001	< 0.0001	< 0.005	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
B-A variability (%)	*±94.3*	*±59.2*	*±78.6*	*±79.2*	*±87.7*	*±102.9*	*±82.6*	*±62.5*	*±52.7*	*±123.6*	*±60.5*	*±52.4*
B-A variability with correction (%)	*±101.3*	*±60.9*	*±84.2*	*±83.2*	*±93.8*	*±96.4*	*±90.8*	*±70.6*	*±51.3*	*±135*	*±60.5*	*±50.8*

* THR, total hip replacement; CORIN, McMinn Hybrid resurfacing; BHR, Birmingham hip resurfacing

sis at our centre in patients at different stages before and after metal-on-metal replacement. They include pre-operative patients, those with unilateral and bilateral hip resurfacings and replacements through to ten years after operation and those with failed metal-on-metal replacement. Between September 2003 and September 2005, we obtained 262 concurrent serum and whole blood specimens from these patients. All were included in the study in order to get a heterogeneous group, which would ensure a wide range of concentration values and provide an unselected cross-sectional sample that would be representative of patients with metal-on-metal arthroplasties. This group includes smaller more homogeneous subsets depending on the type of replacement, unilateral or bilateral replacement and the duration of follow-up as shown in Table I.

The mean age of the patients from whom these specimens were obtained was 56 years (30.2 to 73.4). Of these 189 were men and 73 women. Within this number, 154 had unilateral replacements, 80 had bilateral procedures and the rest were pre-operative controls or followed revision. The implants used in these patients included hip resurfacings and replacements, with large or small diameter bearings and with conventional or short stems (Smith and Nephew Orthopaedics, Warwick, United Kingdom; Midland Medical Technologies, Birmingham, United Kingdom; Corin Medical Industries, Cirencester, United Kingdom; Sulzer Orthopaedics, Winterthur, Switzerland).

The BD Vacutainer system was used in all cases (Beckton Dickinson Medical Pharmaceutical Systems, Oxford, United Kingdom). Whole blood specimens were drawn into two 6 ml lithium heparin Vacutainer tubes and stored at -18°C. In order to obtain serum, blood was drawn into a plain Vacutainer tube and centrifuged almost immediately at 5000 rpm for ten minutes and the supernatant stored frozen at -18°C in two separate 2 ml microtubes (Sarstedt Ltd., Leicester, United Kingdom). One of each frozen tube was kept as a reserve and the other forwarded to the laboratory for analysis. Any delay in centrifuging introduces a potential source of variability in serum analysis through haemolysis. Hence, a biomedical scientist (HZ), ensured that there was no delay in the centrifugation of any of these specimens. De-ionised water flushed through two unused needles and tubes from each batch was analysed to ensure that there was no trace of metal contamination as previously outlined.[10]

Analysis of metal ions was performed with high resolution inductively coupled plasma mass spectrometry with the reporting limits of 0.06 µg/l for serum cobalt and chromium, and 0.1 µg/l and 0.2 µg/l for whole blood cobalt and chromium respectively. The variability between serum and whole blood levels was initially plotted as a normalised scatter to visualise the actual variability between the two measurements. In this scatter, whole blood measurements were normalised to one and the equivalent serum measurements plotted as a scatter. The relationships between serum and whole blood levels were studied using Student's paired *t*-test and agreement was tested using the Bland and Altman limits of agreement.[11] R language was used for statistical analyses.[12] A p-value of ≤ 0.05 was considered significant and all confidence intervals (CI) are quoted at the 95% level. Where regression was used, a simple linear model of the form blood = (α+βserum+ε) was applied. Analysis of the regression residuals for influential data was performed using standard methods.

Results

In 16 specimens the whole blood cobalt levels were below the reporting limit of 0.1 µg/l and in 14 specimens the chro-

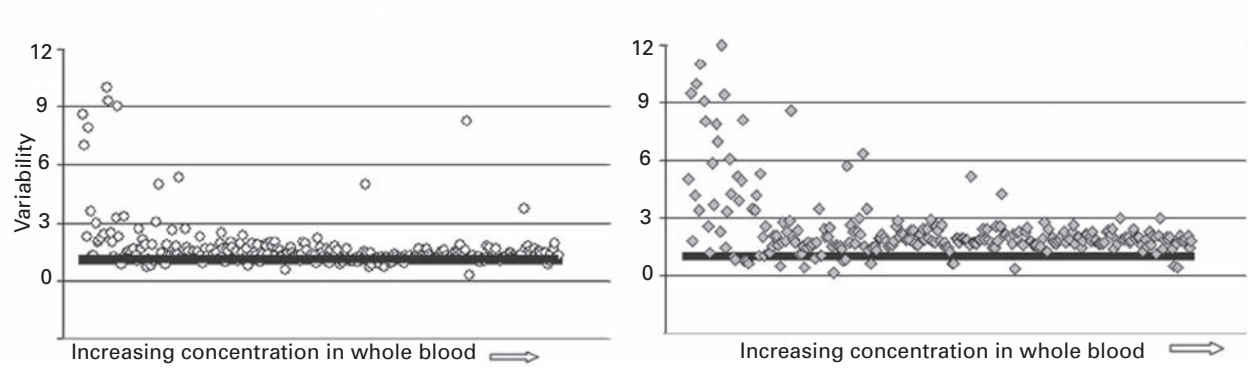

Fig. 1a

Fig. 1b

Scattergraphs showing variability of concentration of a) cobalt and b) chromium in serum per unit concentration in whole blood. The scatter demonstrates that the variability between serum and whole blood is not uniform throughout the range of measurements. At higher concentrations (i.e. on the right of the scale on the X-axes, blood concentrations tending towards > 20) there is less variability, but at lower concentrations (left of the scale, blood concentrations tending towards 0) the variability is greater. The scale on the Y-axes are truncated at the values shown for better appreciation of the detail. Extreme outliers lie beyond the range shown. Their absence does not influence the data displayed.

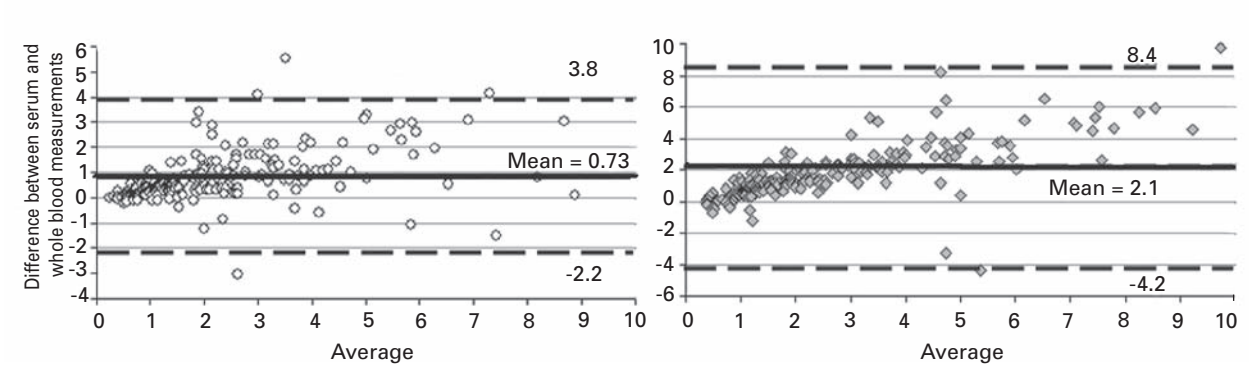

Fig. 2a

Fig. 2b

Scattergraphs showing Bland and Altman limits of agreement between measurements in serum and whole blood for a) cobalt and b) chromium). Only values above the reporting limits have been included (> 0.1 μg/l for cobalt, n = 246 and > 0.2 μg/l for chromium, n = 248). Amongst these six cobalt data points and seven chromium data points lie outside the range displayed but are included in the calculation. The displayed range has been truncated for better appreciation of the detail. Expressed as a percentage of the averages, the limits of agreement for cobalt are -35.2% to +97% and those for chromium are -25% to +144.2%.

mium levels were below the reporting limit of 0.2 μg/l. These were excluded leaving 246 specimens of cobalt concentrations and 248 specimens of chromium for study.

The mean difference between serum and whole blood concentrations of cobalt was 0.84 μg/1 (-3 to 17.2), and of chromium was 2.1 μg/l (-4.4 to 22.7). The differences for both were statistically significant (*t*-test, $p < 0.0001$). The normalised scatter (Fig. 1) showed that the variability was greater in the lower range of concentrations compared with the higher ranges. Bland and Altman analyses (Fig. 2) show the limits of agreement were 3.8 μg/l to 4.2 μg/l and 8.4 μg/l to -4.2 μg/l for cobalt and chromium respectively. Considered as a percentage of the mean, the limits of agreement were 97% and -35.2% for cobalt and 144.2% and -25% for chromium. The difference plots of Bland and Altman (Fig. 2) suggest that there is a trend which is concentration dependent; therefore a linear regression model was fitted to the data (Fig. 3). For the cobalt data, a linear regression model was fitted giving coefficients of $\alpha = -0.36$ (95% CI -0.53 to -0.18) and $\beta = 0.86$ (95% CI 0.83 to 0.88). After an analysis of the residuals ten points were found to be influential. Following their removal the coefficients became $\alpha = 0.01$ (95% CI = -0.10 to 0.13) and $\beta = 0.71$ (95 CI = 0.67 to 0.75). Similarly, the model gave coefficients of $\alpha = -0.12$ (95% CI, -0.26 to 0.01) and $\beta = 0.54$ (95 CI, 0.52 to 0.56) for chromium. An analysis of the residuals suggested 16 measurements were influential. If these are removed the coefficients then become $\alpha = 0.08$ (95% CI, -0.02 to 0.17) and $\beta = 0.48$ (95% CI, 0.46 to 0.51).

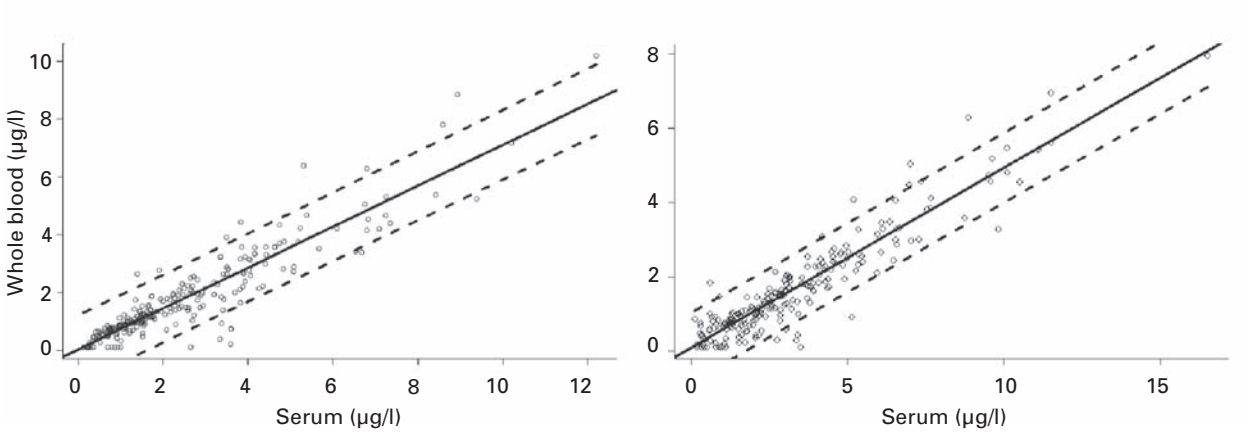

Fig. 3a

Fig. 3b

Scattergraphs showing regression analysis of a) cobalt and b) chromium concentration in whole blood and serum. The solid line represents the regression of whole blood on serum. The broken lines represent the confidence limits for prediction. The potentially over-influential observations have been removed.

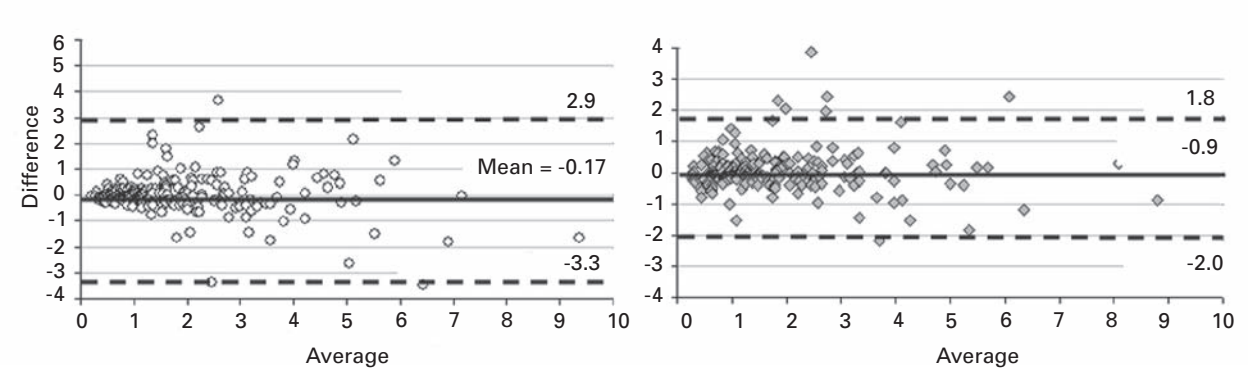

Fig. 4a

Fig. 4b

Scattergraphs showing Bland and Altman limits of agreement for the serum and concurrent whole blood values as in Figure 2, but with correction factors obtained from a linear regression model of the data applied for a) cobalt and b) chromium. These factors are 0.71 for cobalt and 0.48 for chromium (five cobalt data points and seven chromium data points lie outside the displayed range but are included in the calculation). It is obvious that applying these correction factors reduces the mean difference and brings it closer to zero but does not influence the interval between the limits of agreement showing that the degree of variability is not affected.

Applying the correction factors ($\beta = 0.71$ cobalt and $\beta = 0.48$ chromium) obtained from the adjusted regression analyses reduces the mean difference between serum and whole blood concentrations of cobalt and chromium (Figs 4 and 5) and positions the limits of agreement more symmetrically on either side of zero. However their application does not reduce the width of the interval between the limits of agreement (-73% to 67% for cobalt and -94% to 85% for chromium). Analysis of the limits of agreement in smaller more homogeneous subsets (Table I) also does not reduce the variability to an acceptable range.

Discussion

The concentration of metal ions in the blood can be interpreted as a measure of systemic exposure to them. There is great variability in the distribution of metal ions between serum and blood cells. However, in the past, serum metal ion analysis has been performed with graphite furnace atomic absorption spectrometry because of the difficulty in analysing whole blood. With the advances made in high resolution inductively coupled plasma mass spectrometry, poor instrument sensitivity is no longer a problem. Not only is it more sensitive, it is also able to overcome the interference caused by the complex matrix in whole blood[13] making it more reliable.

The unreliability of serum measured with atomic absorption spectrometry is highlighted in a metal ion study[14] comparing 28 mm metal-on-metal total hip replacements (THRs) with ceramic against polyethylene THRs. Serum cobalt levels were assessed with atomic absorption spec-

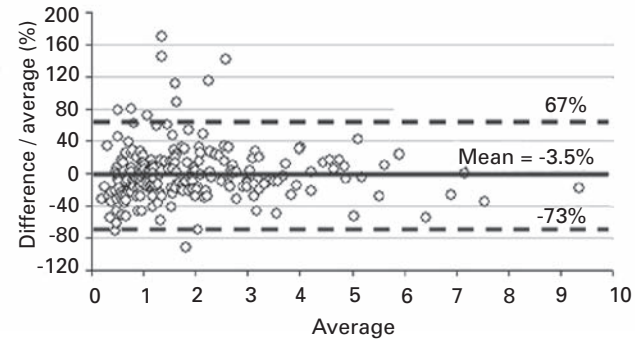

Fig. 5a

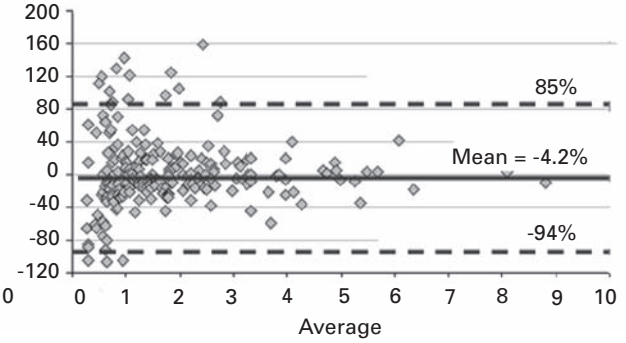

Fig. 5b

Scattergraphs showing Bland and Altman limits of agreement for the same data points as in Figure 4, expressed as a percentage of the average for a) cobalt and b) chromium. This gives an estimate of the varibility as a function of the mean difference. Six cobalt data points and ten chromium data points lie outside the displayed range but have been included in the calculation.

trometry and nearly two-thirds of the specimens were below the limit of detection and had to be given an arbitrary figure (half the limit of detection).

The clinical importance of the unreliability of serum analysis with graphite furnace atomic absorption spectrometry is brought into sharp focus in a second study,[15] which considered the pivotal issue of transplacental metal transfer in pregnant women with hip replacements. This study, in which serum levels were used, led to the conclusion that metal ions do not cross the placenta, which is contrary to current evidence obtained from whole blood analysis using high resolution inductively coupled plasma mass spectrometry.[16] The completely different results question the appropriateness of serum measurements as a surrogate measure of systemic metal levels.

It may still be justified to use serum measurements if the two results are interchangeable or at least interconvertible through the full range of clinical measurements. A surrogate measure can be used interchangeably with a known complete measure if the readings obtained with the surrogate approximate closely with those obtained with the existing measure. If the data do not approximate closely but can be made to conform to the existing measure after the application of a correction factor, it may not be interchangeable but can still be used interconvertibly. If such an agreement is not possible even after the application of appropriate correction factors the value of such a surrogate measure is doubtful.

Interchangeability. In order to use serum and whole blood measurements interchangeably, the mean difference between adequately powered homogeneous or heterogeneous groups of readings should not be statistically significant, and more importantly, the differences between individual readings obtained with the two specimens should not be large. Both cobalt and chromium fail on the paired mean differences. The mean differences were 0.84 μg/l and 2.1 μg/l for cobalt and chromium respectively in the overall heterogeneous group and both were statistically highly significant (*t*-test $p < 0.0001$). The mean differences between readings obtained with the two specimens were statistically significant even in smaller, more homogeneous subsets (Table I).

A significant mean difference in itself does not necessarily rule out the suitability of serum as a surrogate measure, provided a close enough agreement exists between whole blood and serum levels so that one can be used to predict the other. The normalised scatter (Fig. 1) and the Bland and Altman limits of agreement visually illustrate the differences between the individual readings obtained with the two measures (serum and whole blood).

The normalised scatter is essentially a ratio of metal levels in concurrent specimens of serum and whole blood and is a simple method of displaying the variability between individual readings in the two specimens before subjecting the data to statistical analysis. Ratios have been employed in the past by others to highlight the variability of metal ion levels between the different compartments of blood[8] in *in vivo* studies and between cells and supernatant in *in vitro* studies.[17] The scatter (Fig. 1) shows that the variability between individual readings with the two different specimens is particularly large at the lower range, both for cobalt and chromium. This lower part of the range is critical to distinguish a low wear bearing from a high wear bearing.

Bland and Altman[11] propose a simple visual approach based on quantifying the differences found between readings obtained from two different measurements (serum and whole blood in our study). Their method consists of a scatter of data points which represent the differences between individual readings obtained with the two specimens plotted against the average of those two readings. The mean of these data points is shown by a horizontal 'line of agree-

ment' drawn across the graph (solid line in Figs 2, 4 and 5). There are two other parallel lines (interrupted lines in Figs 2, 4 and 5) which are drawn at ± 1.96 times the standard deviations above and below the first line and are termed the 'limits of agreement'. The interval between these two lines contains 95% of the differences between the two methods of measurement under comparison and is a measure of the degree of disagreement between the readings. If it can be judged that the interval between these two boundaries is not clinically important, we can use readings obtained from the two specimens interchangeably.

The only criticism that has been levelled against the Bland and Altman method[18] is that it depends on an expert to make a judgement whether the range of differences evident in their analysis is acceptable or not. However this criticism could equally be applied to every other method of agreement measurement. The lines of agreement were 0.73 μg/l and 2.1 μg/l for cobalt and chromium respectively and the disagreement for both metal ions by the Bland and Altman technique was too wide to be acceptable (Fig. 2). In the present study in order to allow for the different units of measurements (μg/l, nmol/l etc) employed by different laboratories, we plotted the differences as a percentage of the means as well. The percentage limits of agreement were -35.2% to 97% for cobalt and -25% to 144.2% for chromium.

The significant mean differences between groups of patients and the differences between individual readings as seen from the Bland-Altman test, show that readings obtained from serum and whole blood should not be used interchangeably.

Interconvertibility (interchangeability with a conversion factor). Regression analysis provided correction factors for cobalt and chromium. If after application of these factors, the existing mean difference can be reduced and the limits of agreement can be narrowed down to an acceptable level, whole blood and serum levels can be used interconvertibly and serum can still be used as a surrogate measure of metal exposure. Application of correction factors did reduce the mean difference to approach zero (-3.5% for cobalt and -4.2% for chromium) (Figs 4 and 5). However, this linear correction factor obviously does not influence the variability and the limits of agreement exceed a range of ± 65% for cobalt and ± 85% for chromium (Fig. 5) and remains too wide to accept serum and whole blood analyses as interconvertible measurements. The limits of agreement do not reduce to acceptable levels even when the heterogeneous group is subdivided into smaller more homogeneous groups (Table I). Furthermore, the variability shows large differences between each of the individual subsets in the different groups adding an additional element of uncertainty. This suggests that the hypotheses of agreement for both metal ions is rejected on the basis of the limits of agreement of Bland and Altman.

Our results show that the measurement of serum and whole blood metal ion concentrations should not be used interchangeably or interconvertibly. This finding calls into question the appropriateness of using serum as a surrogate measure of systemic exposure to metal ions and the validity of such studies in hip replacement.

Although none of the authors has received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article, benefits have been or will be received but will be directed solely to a research fund, foundation, educational institution, or other nonprofit organisation with which one or more of the authors are associated.

References

1. **Schmalzried TP, Shepherd EF, Dorey FJ, et al.** Wear is a function of use, not time. *Clin Orthop* 2000;381:36-46.
2. **Malchau H, Herberts P, Soderman P, Oden A.** Update and validation of results from the swedish hip arthroplasty registry 1979-1998. Technical report. http://www.jru.orthop.gu.se (date last accessed 26 January 2007).
3. **Kilgus DJ, Dorey FJ, Finerman GA, Amstutz HC.** Patient activity, sports participation, and impact loading on the durability of cemented total hip replacements. *Clin Orthop* 1991;269:25-31.
4. **Daniel J, Pynsent PB, McMinn DJW.** Metal-on-metal resurfacing of the hip in patients under the age of 55 years with osteoarthritis. *J Bone Joint Surg [Br]* 2004;86-B:177-84.
5. **Treacy RBC, McBryde CW, Pynsent PB.** Birmingham hip resurfacing arthroplasty: a minimum follow-up of five years. *J Bone Joint Surg [Br]* 2005;87-B:167-70.
6. **Back DL, Dalziel R, Young D, Shimmin A.** Early results of primary Birmingham hip resurfacings: an independent prospective study of the first 230 hips. *J Bone Joint Surg [Br]* 2005;87-B:324-9.
7. **Keegan GM, Learmonth ID, Case CP.** Orthopaedic metals and their potential toxicity in the arthroplasty patient: a review of current knowledge and future strategies. *J Bone Joint Surg [Br]* 2007;89-B:567-73.
8. **Merritt K, Brown SA.** Release of hexavalent chromium from corrosion of stainless steel and cobalt-chrome alloys. *J Biomed Mater Res* 1995;29:627-33.
9. **MacDonald SJ, Brodner W, Jacobs JJ.** A consensus paper on metal ions in metal-on-metal hip arthroplasties. *J Arthroplasty* 2004;19(8 Suppl 3):12-16.
10. **Daniel J, Ziaee H, Salama A, Pradhan C, McMinn DJW.** The effect of the diameter of metal-on-metal bearings on systemic exposure to cobalt and chromium. *J Bone Joint Surg [Br]* 2006;88-B:443-8.
11. **Bland JM, Altman DG.** Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;8476:307-10.
12. **No authors listed.** R Development Core Team R. *A language and environment for statistical computing*. http://www.r-project.org/ (date last accessed 26 January 2007)..
13. **Case CP, Ellis L, Turner JC, Fairman B.** Development of a routine method for the determination of trace metals in whole blood by magnetic sector inductively coupled plasma mass spectrometry with particular relevance to patients with total hip and knee arthroplasty. *Clin Chem* 2001;47:275-80.
14. **Brodner W, Bitzan P, Meisinger V, et al.** Elevated serum cobalt with metal-on-metal articulating surfaces. *J Bone Joint Surg [Br]* 1997;79-B:316-21.
15. **Brodner W, Grohs JG, Bancher-Todesca D, et al.** Does the placenta inhibit the passage of chromium and cobalt after metal-on-metal total hip arthroplasty? *J Arthroplasty* 2004;19(8 Suppl 3):102-6.
16. **Ziaee H, Daniel J, Blunt S, Datta A, McMinn DJW.** Transplacental transfer of cobalt and chromium in patients with metal-on-metal hip arthroplasty. *J Bone Joint Surg [Br]* 2007;89-B:301-5.
17. **Merritt K, Wenz L, Brown SA.** Cell association of fretting corrosion products generated in a cell culture. *J Orthop Res* 1991;9:289-96.
18. **Stockl D, Cabaleiro DR, Uytfanghe KV, Thienpont LM.** Interpreting method comparison studies by use of the Bland-Altman plot: reflecting the importance of sample size by incorporating confidence limits and predefined error limits in the graphic. *Clin Chem* 2004;50:2216-18.