

ORIGINAL COMMUNICATION

Total body water measurement using bioelectrical impedance analysis, isotope dilution and total body potassium: a scoring system to facilitate intercomparison

CH Gonzalez^{1,2*}, JA Evans², SW Smye² and P Holland³

¹Department of Physics, University of Caldas, Colombia; ²Department of Medical Physics, University of Leeds, Leeds, UK; and ³Department of Paediatrics, University of Leeds, Leeds, UK

Objectives: (1) To develop a scale that is useful in evaluating the accuracy of multifrequency bioelectrical impedance analysis (MF-BIA) in the assessment of body water volumes against the accepted gold standard measurements based on isotope-dilution and total body potassium (TBK). (2) To perform a pilot test of the scale.

Design: A scale was developed to evaluate the accuracy of MF-BIA in the assessment of body water volumes. Questions were obtained from reading the scientific literature and discussions involving the four authors. Three of these and two additional independent readers pre-tested the scale. A weighting was identified for each question and a pilot test with a sample of 10 articles (different to those used for the questionnaire performance) was conducted. A further validation was carried out with a second set of 20 articles and two additional independent readers.

Results: The kappa statistic expressing the level of agreement between pairs of the first three authors using this scale with 10 articles, was 0.3, 0.4 and 0.6 after the first attempt. A second evaluation after specific changes improved the agreement to 0.8, 0.6 and 0.8. The mean score for 10 articles was 252 ± 36 points from a total score of 400 ($63 \pm 9\%$). The evaluation with the second set of 20 articles resulted in a κ of 0.7 from two pairs of authors. The evaluation with two additional reviewers resulted in a $\kappa = 0.7$.

Conclusion: A tool has been developed to assess the accuracy of the MF-BIA technique and to identify methodological components, plan future studies and critically evaluate data in this area. It is likely that this tool may also be used to assess the accuracy of single frequency studies.

Sponsorship: COLCIENCIAS and University of Caldas, Colombia sponsor CH Gonzalez.

European Journal of Clinical Nutrition (2002) **56**, 326–337. DOI: 10.1038/sj/ejcn/1601316

Keywords: bioelectrical impedance; dilution techniques; TBK; body water; scoring system

Introduction

A major challenge in clinical practice is to accurately monitor acute and chronic changes in hydration during various diseases and their clinical management (Johnson *et al*, 1996; Pullicino *et al*, 1992). Thus, the development of methods for measuring and monitoring the volume of body water and its distribution has been of increasing recent interest (Ellis &

Wong, 1998). Methods based on bioelectrical impedance analysis (BIA) or its multifrequency variant (MF-BIA) have attracted attention because of their low cost and high convenience. The validation of these techniques normally involves using isotope dilution (Brummer *et al*, 1992) and total body potassium (TBK) as gold standards (De Lorenzo *et al*, 1995). There has been a rapid proliferation of BIA literature. A search from 1985 to 1990 with the keyword: 'bioelectrical impedance' (when Lukaski's article appeared), retrieved 73 studies; from 1991 to 1995 there were 271 articles and, from 1996 to 2001, 523 articles. This was only in Medline database. However there are conflicting and confusing results and conclusions about its accuracy and its utility (Schoeller, 1996a). Variability in the quality of BIA studies may have contributed to these inconclusive and

*Correspondence: CH Gonzalez, c/o JA Evans, Department of Medical Physics, Leeds General Infirmary, Gt George Street, Leeds LS1 3EX, UK. E-mail: labie@cumanday.ucaldas.edu.co

Guarantor: CH Gonzalez.

Contributors: JAE, SWS and PH were all involved in study conception, design, interpretation of results and approving final manuscript.

Received 14 November 2000; revised 23 July 2001; accepted 25 July 2001

conflicting findings. Differences of quality across the studies may indicate that the results of some are more reliable than those of others. Low quality studies are more likely to produce incorrect inferences and conclusions regarding the accuracy and usefulness of BIA. Therefore, evaluation of BIA should be based on evidence from high-quality studies.

The purpose of this paper was to describe a scale to evaluate the quality of BIA studies and to define and discuss the evaluation criteria which might be more generally used.

Scale development methods

The process used to perform our study consisted of eight steps shown in Figure 1. Steps 1–7 are preparation with the final evaluation (step 8) following at the end. The details of steps 1–7 are described below.

1. Literature review to identify key characteristics to be extracted from each paper through examination of:

- a. Key characteristics identified as being paramount in published checklists evaluating quality of studies published

in the health technology assessment literature to identify general headings for evaluation. (Bland *et al*, 1985; Cho & Bero, 1994; Detsky *et al*, 1992; Khan *et al*, 1996; Moher *et al*, 1995; 1996; NHS Centre for Reviews and Dissemination, 1996; Oxman, 1994)

- b. Possible bias (sampling, diagnostic access, spectrum, co-intervention, verification, withdrawal, work-up, incorporation disease progression, observer variability, diagnostic review, test review, clinical review and loss to follow-up bias (Kelly *et al*, 1997).
- c. Articles about BIA, TBK and dilution techniques standardisation (Deurenberg *et al*, 1988; Kushner *et al*, 1996; National Institutes of Health Technology Assessment Conference Statement, 1994; Schoeller *et al*, 1985; Schoeller, 1996b; Lykken *et al*, 1983).
- d. Articles about the use of statistics in scientific literature (Altman, 1980, 1991; Altman *et al*, 1983; Emerson & Colditz, 1983; Gardner *et al*, 1983; 1986; Jenicek, 1989).
- e. Views expressed by the authors who have some knowledge and experience in the area: two medical physicists, one paediatrician and one physician.

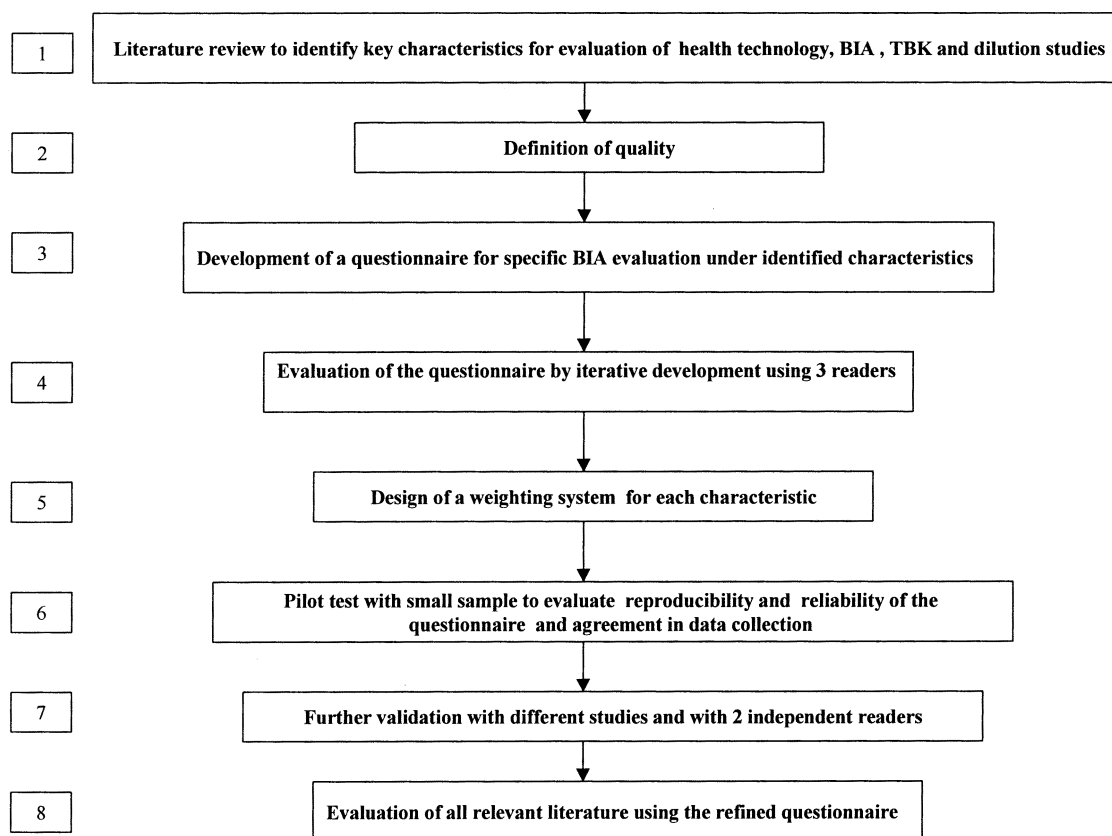


Figure 1 Scale development process.

This process resulted in steps 2 and 3 below

2. Definition of quality

Quality as defined for our purposes of evaluating BIA estimates of fluid volumes was based on the following subheadings:

- a. Study design.
- b. Bias.
- c. Methodological performance.
- d. Statistical analysis.

3. Development of a questionnaire for specific BIA evaluation

Under the headings identified above, a questionnaire was composed consisting of 66 questions. Additionally a form was developed to extract the data from the articles.

4. Evaluation of the questionnaire by iterative development using three independent observers

Three authors with different backgrounds (two medical physicists and one physician studying the BIA technique) pre-tested the questionnaire three times using 10 articles in order to decide which questions should remain in the final questionnaire. Specific suggestions were:

- a. To record a 'No' when a procedure was specifically mentioned as not done but 'Not clear' if the procedure was mentioned but without sufficient detail to allow replication.
- b. To split compound questions in order to achieve more specificity.
- c. To eliminate duplication in the questions.

The final questionnaire consists of 21 explicitly defined questions (Appendix 1). Questions 12 and 13 refer to three sub-questionnaires about procedures on BIA, dilution and TBK techniques respectively. Possible responses are: 'Yes', 'No', 'Not clear', or 'Not applicable'.

The questionnaire was divided into four main sections:

- a. Section A examines the design of the study and inclusion/exclusion criteria for its subjects.
- b. Section B looks for description of the subjects and how they were selected.
- c. Section C is related to the procedures for carrying out the BIA and the gold standard techniques.
- d. Section D concerns calculations and statistical analysis.

Forty-two percent of questions are related only to the quality of the research design, 16% to the quality of the report writing and structure and 42% consider both topics. The original version of the questionnaire was modified to make it simpler and more self-explanatory.

5. Identify weighting for each characteristic

A theoretical weighting scheme was used to score the studies according to Streiner and Norman (1995). In this case the weighting scheme was established from an analysis of the literature and from local experimentation. The following factors were given heavy weighting: study design; sample selection; BIA; TBK and dilution procedures, hydration status; and statistical analysis. Each section was scored on a continuous scale from 0 to 100 for positive answers. Within a section each item was weighted according to its impact on results when data were available in the scientific literature or when a variable could influence the results in more than one technique (references used to give the weight for each variables are at www.medphysics.leeds.ac.uk/~chgc/vienaref.htm). When the data were not available in the literature, questions were weighted equally. Negative answers did not receive a score and the 'Not applicable' option allowed reviewers not to penalise a paper if one item was not necessary for the study. The maximum possible score was 400 points. Figure 2 shows the extent to which each item contributes to the total score for an article.

6. Pilot test

When the reviewers had agreed on the final version of the questionnaire, three evaluations were conducted for the assessment of reproducibility and reliability. An additional test was performed to evaluate agreement in data extraction. For the first evaluation a computerised MEDLINE search was performed to identify articles which used BIA and dilution or TBK techniques and which were published between 1969 and 1998 (inclusive). This covered the period between the original article by Hoffer *et al* (1969) and the time of the study. The exercise was performed as if these articles were to enter to a meta-analysis. The search strategy was exploration by *bioelectrical impedance analysis* (or 17 substitutes) and *body water* (or 13 substitutes) or dilution (or nine substitutes) and *TBK* from 1969 to 1998. The search was limited to studies with human beings, English and Spanish languages. Letters, abstracts and editorials were also excluded. The initial search yielded 161 citations.

One of the reviewers read the abstracts and selected those that had used dilution techniques (TBW/ECW) and/or total body potassium (ICW) as gold standard. Seventy-two studies were eligible. From these, random numbers generated by a calculator selected five articles and five additional articles were chosen for the pilot test to give a variety of ages and clinical conditions.

Three reviewers tested the final questionnaire. The revision was made on the basis of inter-rater agreement and reviewers' comments on content validity of individual questions. The reviewers worked independently. To minimise observer bias, they were given masked photocopies of the 10 articles. The methods sections of these studies were photocopied in such a way that the author and journal names, dates, and all other reference information had been

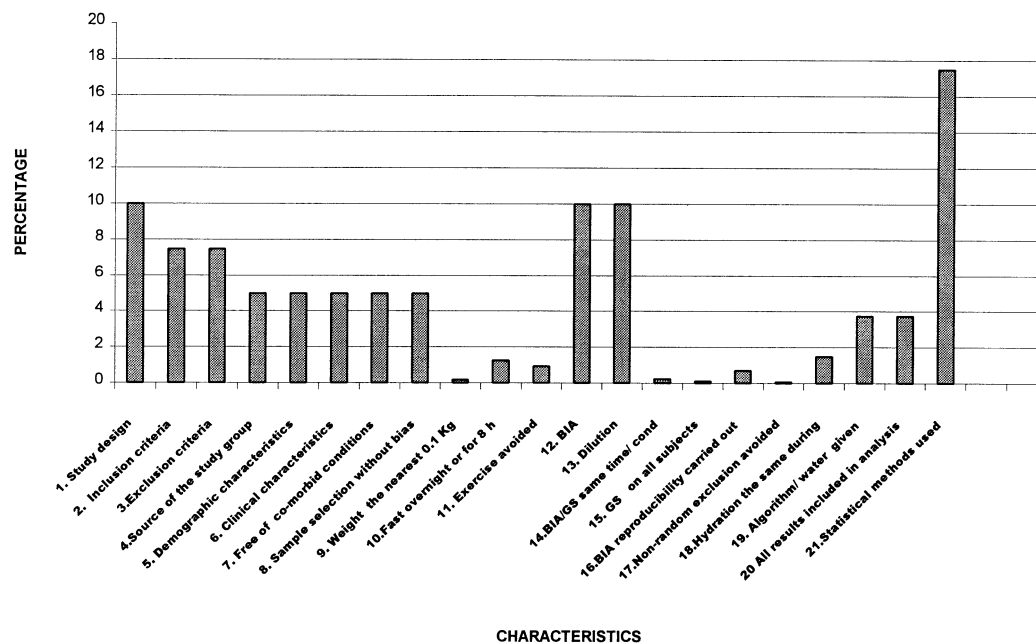


Figure 2 Contribution (%) of characteristics to score of MF-BIA/dilution studies.

omitted. General guidelines were given on the cover sheet for the use of the scale. An additional questionnaire to describe the areas of particular strength, reason for doubt or for rejection, time spent with the evaluation of each study, and the influence of the knowledge of one of the articles on the final decision on it was attached to each paper.

The first evaluation of reproducibility and reliability produced results which were poor, as expressed by the kappa statistic ($\kappa < 0.6$). A second evaluation was performed after making the following changes:

- The low agreement in the first evaluation was generally the result of discrepancies in the use of the response 'Not applicable' and probably fast reading of the articles. Therefore, the results of the questionnaires were corrected when the answers were unambiguous (side of the body on which the measurements were performed, baseline samples for dilution techniques, etc).
- The instruction manual was expanded to explain in more detail definitions of terms and the importance of each question. At this time, the reviewers were given a six-page set of written instructions.
- Reviewers suggested that when references were cited in the 'Methods' section, these references should be included with the article. For the second evaluation the relevant bibliography was therefore attached to each article.

The kappa statistic for the same 10 articles was improved after these changes ($\kappa = 0.6$). An evaluation was done with a form developed for data extraction. Three reviewers tested the form by using three of the 10 articles (results not shown).

Selection of the articles. The acceptability of each article was described as 'Definitely yes', 'Probably yes' or 'Not acceptable' on the basis of the score. Differences among reviewers were resolved by re-examining the report and discussing the particular point.

Inclusion criteria. Studies were selected after three assessments by unanimous decision or by agreement between two of the three reviewers: studies with a score $\geq 60\%$ of the total possible score was qualified as 'Definitely yes'; studies that were scored from $\geq 50-59\%$ of the total possible score were qualified as 'Probably yes'; and the final inclusion or exclusion of these articles was made by a third evaluation between the readers.

Exclusion criteria. Studies with a score of $< 50\%$ of total possible score or studies classified as 'Probably yes' which after a third evaluation still had an score $< 60\%$.

7. Further scale validation

To test whether the repeated review of the same 10 articles had influenced the results, a second different set of 10 masked articles was reviewed independently by two pairs of the reviewers (20 different articles in total).

After a consistent agreement was obtained between us, two independent reviewers with experience in the bioimpedance area but not involved in the development of the instrument were asked to qualify three randomly selected and masked articles.

Table 1 Scale to evaluate studies which use bioelectrical impedance analysis compared to dilution and TBK techniques in estimation of body water volumes

Variable	Subdivisions	Score subtotal	Definition/comment	Reason of weighting
A. Study design	1. What type of study design was used?		(Chalmers et al, 1981)	Cho and Bero (1994) assign different weight to these three types of studies in ascendant order of quality. However, because the scale was designed to evaluate BIA compared to dilution and TBK measurements and not interventions on the subjects, uncontrolled and cohort or case control were given similar weight.
	Case report	5		
	Uncontrolled study	30		
	Cohort or case control	40	Because the study is prospective and data recording can be controlled, Cohort study is the best design in observational studies	
	Not clear			
2. Were the inclusion criteria clarified?		30	To avoid sampling bias and popularity bias. Arbitrary recruitment of subjects could invalidate the results of a study (Cho & Bero, 1994; Altman et al, 1983; Gardner et al, 1986; Baar & Tannock 1989; Bailar & Mosteller, 1988; Chalmers et al, 1981; Mulrow et al, 1989)	Because inclusion/exclusion criteria can effect the result with the same bias they were weighted equally
3. Were the exclusion criteria clarified?		30	To avoid sampling bias. When eligible subjects are not admitted to the study for some reason, they should be described to indicate if they differ from those included in the study (Chalmers et al, 1981; Evans & Pollock, 1985; Reisch et al, 1989)	
Subtotal		100		
B. Subjects	4. Were the recruitment methods of the study group described?	20	To avoid sampling bias. Useful to compare different studies and for generalization of the study results. (Bland et al, 1985; Altman et al, 1983; Gardner et al, 1986; Mulrow et al, 1989; Goodman et al, 1994)	Because all items in this category can cause the same bias and affect the results they were weighted equally
	5. Was there a description of the demographic characteristics of the subjects? (race, gender, age, height, BMI)	20	To avoid sampling bias. To evaluate if they are typical for subjects for whom the study was intended (Cho & Bero, 1994; Altman et al, 1983; Lionel & Herxheimer, 1970; Mulrow et al, 1989; Working group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature, 1994)	
6. Was there a description of the clinical characteristics of the subjects? (normal or type) of disease and severity)		20	To avoid sampling bias. To compare results from similar studies (Altman et al, 1983; Mulrow et al 1989).	
7. Was the study group free of co-morbid conditions?		20	If there were other pathologies different to those studied, the results can be altered (Mulrow et al, 1989; Goodman et al, 1994).	
8. Was the sample selected in a way that avoids bias?		20	Sampling bias (Kelly et al, 1997)	
Subtotal		100		
C. Procedures	9. Was the weight measured to the nearest 0.1 kg? (1 – 5 g in children under 3 y)	0.7	This variable can affect results in both BIA and dilution techniques	Error of 1 kg in weight results in an error of 0.2 l in predicted TBW by BIA (Kushner et al, 1996) a similar effect caused by changing the electrode size
	Subtotal			

Table 1 continued.

Table 1 Continued

Variable	Subdivisions	Score subtotal	Definition/comment	Reason of weighting
C. Procedures				
10. Did the subjects fast overnight or for 8 h before the test? (3 h before in children < 3 y)		5	This variable can affect results in both BIA and dilution techniques	Studies have shown that this variable can effect the body resistance about 2.9% (Gonzalez et al, 1999), which is slightly above of the effect caused by limb abduction
11. Did the subjects avoid strenuous exercise for 8 h before the test?		3.7	Strenuous exercise affects BIA, dilution and TBK results (Schoeller, 1996b; Deurenberg et al, 1988; Lykken et al, 1983).	Studies have shown that this variable can effect the body resistance about 2.1% (Gonzalez et al, 1999), which is similar to the effect caused by changes in electrode position
12. Was the BIA performed in a standardised manner?	Go to Table 1a	40	(Lionel & Herxheimer, 1970)	Because in the future we intend to assess BIA accuracy, procedures were weighted higher
13. Was the gold standard (GS) test performed in a standardised manner?	Go to Table 1b and 1c	40	(Lionel & Herxheimer, 1970)	Because in the future we intend to assess BIA accuracy, procedures were weighted higher
14. Were BIA and GS tests done at the same time and in similar conditions?		1		Body water turnover is rapid and measuring at different times can alter the results however the impact of this variable has not been quantified. An arbitrary value was assigned
15. Was the GS performed on all subjects?		0.5	To avoid Verification bias (Kelly et al, 1997; Altman et al, 1983; Mulrow et al, 1989).	An arbitrary value was assigned
16. Was a test of BIA reproducibility carried out?		2.8	(Mulrow et al, 1989; McMaster University Health Sciences Centre, 1981)	The coefficient of variation within day and between day is 1 – 2% (Kushner et al, 1986). These are similar to errors caused by the effect of bladder status and electrode position. The mean value of these variables was assigned in this case
17. Was non-random exclusion of subjects avoided?		0.3	To avoid withdrawal bias (Kelly et al, 1997; Chalmers et al, 1981; Reisch et al, 1989)	An arbitrary value was assigned
18. Was the hydration status likely to have remained the same during the whole measurement period?		6	BIA and dilution techniques require that the hydration status during the time of measurements stay almost constant to obtain reliable results	Studies have shown that this variable can affect the body resistance about 6.4% (Streiner & Norman, 1995). This value is between the effect caused by changing the skin temperature and posture. A value approximately equal to the mean of them was assigned in this case
Subtotal		100		
D. Calculations and statistical analysis				
19. Was the algorithm to water estimation given?		15	Different algorithms give different results (Cox & Soeters, 2000)	Using linear regression or Hanai analysis can change estimates of TBW between 0.75 and 2.75% respectively and 4.1 – 5.6% for ECW respectively (Cox & Soeters, 2000)
20. Were all results included in analysis?		15	To control selection bias, results should be informed with any without inconsistent data (Cho & Bero, 1994; Chalmers et al, 1990)	The impact of this variable on the final estimates of body water is unknown. An arbitrary value was assigned
21. What was the statistical analysis used?			(Cho & Bero; Chalmers et al, 1981; Working group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature, 1994)	Authors agree that this component is very important in assessing studies quality, then this variable was weighted higher
	Parametric test	30		
	Correlation/regression	50		
	Bland/Altman	70	(Altman et al, 1983; Bland & Altman, 1986)	Authors recognise nowadays that Bland/Altman is a better approach to compare two methods
Subtotal		100		
Total		400		

How to use the scale. After revision of the study, each question was answered by ticking the appropriate box in the form. Only questions with positive or 'Not applicable' answers were weighted. When the study compared BIA with both dilution and TBK techniques, the weights applied to these questions were added to the gold standard score.

Statistical analysis

Percentage agreement and kappa statistic were used for the analysis of agreement between reviewers on study choice. Acceptable agreement was set at a kappa level from 0.6 to 1.0. Kappa interpretation was made according to Altman (1991). The score of the articles is expressed as a percentage of the total possible score.

Results

Three reviewers who used the quantitative scale to rank independently the set of 10 articles evaluated reproducibility, reliability and applicability. There were only three instances of missing data: two from one reader and one for another reader. Table 1 shows the final questionnaire with the references on which the score was based and Appendix 1 the forms used for each article. The test required approximately 20 min per article per reviewer. Table 2 shows the kappa values and percentage inter-rater agreement between

reviewer pairs for their summary ranking scores on the same 10 articles in two occasions. The final result of the evaluation of the articles is given in Table 3. After the 'Probably yes' articles were reviewed again among the three reviewers. The mean of reviewer scores for 10 articles was 251.6 ± 35.9 points from a total score of 400 ($62.9 \pm 8.98\%$; Table 4). In the further validation of the scale between two pairs of reviewers with another set of 10 articles per pair, the kappa statistic was 0.7 (CI 0.6–0.7) and 0.7 (CI 0.6–1.3).

The agreement between the two independent reviewers was $\kappa=0.7$ (confidence interval 0.6–0.8) without corrections for solving differences. Altman (1991) considers this agreement as good. The average time for new users of the scale (including the manual) was 17 min.

Discussion

In order to perform a systematic review of primary research of variable quality, it is necessary to have an adequate assessment of study quality which guarantees that the results are likely to be free from bias and consequently can be generalisable (Khan *et al*, 1996).

We developed a standardised scale that could be used to score the quality of BIA studies in comparison with gold standard techniques. The scale can be used as a simple checklist or to provide a quantitative summary score. It does not apply uniform weighting across all questions. It weights them

Table 1a Question no. 12 of the questionnaire to evaluate studies BIA procedures

Variable	Score	Reason
1. Was the bladder emptied before test?	1.8	Studies have shown that this variable can affect the body resistance about 1.0% (Gonzalez <i>et al</i> , 1999) ^a
2. Was the menstrual cycle considered or not applicable?	2.1	Studies have shown that this variable can affect the body resistance about 1.2% (Gonzalez <i>et al</i> , 1999)
3. Was a non-conductive surface used for the test?	1.1	Studies have shown that this variable can affect the body resistance about 0.6% (Gonzalez <i>et al</i> , 1999)
4. Was the skin cleaned with alcohol?	1.0	Studies have shown that this variable can affect the body resistance about 0.5% (Gonzalez <i>et al</i> , 1999)
5. Was the environmental or skin temperature noted?	13.1	Studies have shown that this variable can affect the body resistance about 7.3% (Gonzalez <i>et al</i> , 1999)
6. Was any metallic object removed from the subjects?	1.1	Studies have shown that this variable can affect the body resistance about 0.6% (Gonzalez <i>et al</i> , 1999)
7. Was the subject position and time in that position recorded?	5.9	Studies have shown that this variable can affect the body resistance about 3.3% (Gonzalez <i>et al</i> , 1999)
8. Was the height measured to the nearest 0.5 cm? (0.1 cm in children under 3 y)	0.6	An over- or under-estimation of 2.5 cm in height leads to an error of ≈ 1 l in TBW estimation (Kushner <i>et al</i> , 1996). This error is only slightly lower than that produced by changing electrode size
9. Was the abduction of limbs described?	4.8	Studies have shown that this variable can affect the body resistance about 2.7% (Gonzalez <i>et al</i> , 1999)
10. Was the electrode position described? (3 cm being the minimal distance between the sense and voltage electrodes in children less than 3 y)	3.7	Studies have shown that this variable can affect the body resistance about 2.1% (Gonzalez <i>et al</i> , 1999)
11. Was the brand or size of electrodes stated?	0.7	Studies have shown that this variable can affect the body resistance about 0.4% (Gonzalez <i>et al</i> , 1999)
12. Did the study state the side of the body on which measurements were taken?	4.1	Studies have shown that this variable can affect the body resistance about 2.3% (Gonzalez <i>et al</i> , 1999)
Subtotal	40	

^aReferences available at: www.medphysics.leeds.ac.uk/~chgc/vienaref.htm

Table 1b Question no. 13 dilution techniques standardisation

Variable	Score		Reason
1. Were drinks avoided for several hours before the test?		(Schoeller, 1996b)	Because it was found in the literature what is the impact of these variables in the results, they were weighted equally
2. Were baseline physiologic samples taken before the dose of tracer?	4.4	(Westertep et al, 1995)	
3. Did each subject take the complete dose? (The container used for the dose should be washed out with water and this water should be given to the subject)	4.4	(Jebb & Elia, 1993; Schoeller et al, 1985)	
4. Was the reported period of equilibration adequate? TBW: 3–4 h ECW: 3–4 h	4.4		
If ECW expanded TBW: 4–5 h ECW: 5–6 h			
6. Was the measurement procedure the same for all subjects?	4.4		
5. Were the samples stored in airtight tubes?	4.4	(Jebb & Elia, 1993)	
7. Was eating/drinking avoided during the equilibration period?	4.4	(Schoeller, 1996b)	
8. If the back-extrapolation method was used were foods avoided for 1 h after the dose or not applicable?	4.4		
9. If urine was sampled for TBW was the first specimen discarded and then two samples collected after the dose or it was not applicable?	4.4	(Jebb & Elia, 1993)	
<i>Subtotal</i>	40		

Table 1c Question no. 13 total body potassium standardisation

1. Did the subjects change their clothes before the test?	13.33	(Forbes, 1987)	Literature does not inform the proportional impact of these variables on the estimates of intracellular water. They were weighted equally
2. Was jewellery removed?	13.33	(Ellis, 1996)	
3. Was the calibration of the instrument with phantoms confirmed?	13.33	(Jebb & Elia, 1993, Ellis, 1996)	
<i>Subtotal</i>	40		

Table 2 Agreement between three reviewers

Reviewers	Agreement (%)	Kappa	Standard error (K) (kappa)	Confidence intervals	Interpretation
<i>First agreement between three reviewers</i>					
Reviewers 2–3	70	0.3	0.36	–0.5–1.0	Fair
Reviewers 1–3	70	0.4	0.27	–0.1–1.0	Moderate
Reviewers 1–2	80	0.6	0.26	0.1–1.0	Moderate/good
<i>Second agreement between three reviewers</i>					
Reviewers 2–3	90	0.8	0.17	0.4–1.1	Good
Reviewers 1–3	80	0.6	0.26	0.4–0.7	Moderate/good
Reviewers 1–2	90	0.8	0.17	0.4–1.1	Good

according to defined criteria. Scoring for each attribute thereby defines its relative contribution to the total score.

We chose '> 60%' as the standard for the highest category of acceptance to perform the pilot test. It was similar to Chalmers *et al* (1990) in which a score of > 6/9 was regarded as good. This was a cohort study of summary reports of

controlled trials. In another study of 242 randomised controlled trials Rochon *et al* (1994) found an average score of $37.2 \pm 13.1\%$. We concluded that the greater of these two criteria was more appropriate.

This scale is an innovation in BIA studies and it is important for several reasons: firstly, because reviewers can

Table 3 Selection of articles

Agreement	Accepted	Probably	Rejected
<i>First evaluation</i>			
Unanimous	2	3	
Two reviewers	3	1	1
Total = 10	5	4	1
<i>Second evaluation</i>			
Unanimous	5	3	
Two reviewers	1	1	
Total = 10	6	4	0
<i>Third evaluation</i>			
Unanimous	6	0	4
Total = 10	6	0	4

Table 4 Final score for 10 articles

Article number	Score	%	Reference
1	213.86	53.47	Hannan et al (1995)
2	198.92	49.73	Pullicino et al (1992)
3	262.6	65.65	Armstrong et al (1997)
4	255.14	63.79	Cornish et al (1996)
5	327.16	81.79	Visser et al (1995)
6	280.14	70.04	Ho et al (1994)
7	259.33	64.83	De Lorenzo et al (1995)
8	252.54	63.14	Van Marken Lichtenbelt et al (1997)
9	238.74	59.69	Johnson et al (1996)
10	228.36	57.09	Deurenberg et al (1996)
Mean	251.68	62.92	
s.d.	35.94	8.98	

use it to identify critical methodological components in this area. Secondly, researchers can use it to plan their studies; using the scale ensures that important factors such as room temperature, which can strongly affect results, will not be forgotten. Thirdly, literature reviewers wishing to summarise information on BIA can use it to evaluate and rank the quality of existing pertinent literature. The scale can help authors and editors to ensure that all the fundamental information needed in the reports is included (Lionel & Herxheimer, 1970).

The scale may be also used as a tool for standardisation of studies that could be candidates for inclusion in meta-analyses. Meta-analysis is a novel scientific technique, which allows a more objective appraisal of the evidence than do traditional narrative reviews. It provides a more precise estimate of the effectiveness of health technology and may explain variability between the results of individual studies (Hardy & Thompson, 1998).

Methodological quality of a study is defined as a minimisation of systematic bias and consistency of conclusions with results (Cho & Bero, 1984). It is possible to determine the methodological quality of a study only to the extent that the study design and analytical methods are reported. Our process measures criteria critical to achieving high methodological quality; however, in consequence, these criteria should be adequately reported in any article using BIA, dilution and TBK techniques. The lack of important items of description in some apparently well-executed studies may

at times be due to the short space authorised by editors for publication.

The ideal kappa for agreement between raters is +1 (Altman, 1991). In the pilot test, the first agreement between the reviewers was below the desirable score (> 0.6). Similar studies report a kappa average greater than 0.7 for all reviewers (Khan et al, 1996). A second assessment of the same 10 articles was made using more detailed instructions and more precise definitions of terms. The kappa for this assessment was within the desirable range. Finally, further discussions and clarification resulted in complete agreement (Table 3).

The scale had good reliability and validity when used by reviewers from different backgrounds who worked completely independently and without lengthy training sessions. Independence of reviewers avoids the systematic bias that can be introduced by consensus conferencing if one of the reviewers has convincing opinions (Cho & Bero, 1994).

We performed a further evaluation of the scale by using a second different set of 20 masked articles the agreement was 'good' according to the Altman (1991) classification. Two additional reviewers that work in the bioimpedance field obtained also a 'good' agreement without any further correction. This reinforces the reliability of the scale.

Any method of evaluating quality of published scientific studies stimulates discussion of validity. The present scale has evolved over a period of 3 y and will probably be changed as experience is gained in this field. Among the possible defects are variation due to subjective judgements, inaccuracies due to incomplete reporting, inclusion of questions that might not be considered meritorious by some experts, and exclusion of some questions considered important by others. To some extent, these problems have been handled by the method of differential photocopies of the studies and by multiple scoring followed by discussion.

We are aware of the influence of the machine in the MF-BIA results. However, we have not made a judgement at this initial stage as to which machine is the most accurate and therefore the identification of the machine does not properly form a part of the quality assessment instrument. Having developed the tool it would then be possible to use it to study the influence of machine type on the accuracy of measurements and this would indeed be a useful, but separate study.

The issue of algorithm selection (regression, mixture theory, etc) is slightly different and we believe that explicit statement of the derived algorithm used to calculate body water is a characteristic of a good quality study. Therefore, in question 19 we address the matter related to the algorithm. The judgement of whether the algorithm used was the correct one or not is beyond the study objectives. In this study, we did not set out to say which one was better, although a separate study using this tool could tackle this issue.

To develop the scale we have made and extended review of literature. However, we would like to discuss with others the performance of the instrument and we are ready to update it when it may be necessary. We will prepare the

scale to be discussed in a web page in the near future (www.medphysics.leeds.ac.uk/~chgc/BIAscoresystem.htm).

This scale is intended for use in a future evaluation of BIA literature compared with dilution and TBK techniques.

References

- Altman DG (1980): Statistics and ethics in medical research. *Br. Med. J.* **28**, 1336–1338.
- Altman DG (1991): *Practical Statistics for Medical Research*. London: Chapman and Hall.
- Altman DG, Gore SM, Gardner MH & Pocock SJ (1983): Statistical guidelines for contributors to medical journals. *Br. Med. J.* **286**, 1489–1493.
- Armstrong LE, Kenefick RW, Castellani JW, Riebe D, Kavouras SA & Kuznicki JT (1997): Bioelectrical impedance spectroscopy technique: intra-, extracellular and total body water. *Med. Sci. Sports Exerc.* **29**, 1657–1663.
- Baar J & Tannock, I (1989): Analysing the same data in two ways: a demonstration model to illustrate the reporting and misreporting of clinical trials. *J. Clin. Oncol.* **7**, 969–978.
- Bailar JC & Mosteller F (1988): Guidelines for statistical reporting in articles for medical journals. *Ann. Intern. Med.* **108**, 266–273.
- Bland JM & Altman DG (1986): Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **8**, 307–310.
- Bland JM, Jones R, Bennet S, Cook DG, Haines AP & Macfarlane AJ (1985): Is the clinical trial evidence about new drugs statistically adequate? *Br. J. Clin. Pharmacol.* **19**, 155–160.
- Brummer RJM, Bengtsson BA & Bosaeus I (1992): Validation of body composition determination by bioelectrical impedance analysis in acromegaly. *Eur. J. Clin. Nutr.* **46**, 47–52.
- Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D & Ambroz A (1981): A method for assessing the quality of a randomised control trial. *Control Clin. Trials* **2**, 31–49.
- Chalmers I, Adams M, Dckersin K, Hetherington J, Tarnow-Moedi Q, Meinert C, Tonascia S & Chalmers TC (1990): A cohort study of summary reports of controlled trials. *J.A.M.A.* **263**, 1401–1405.
- Cho MK & Bero LA (1994): Instruments for assessing the quality of drug studies published in the medical literature. *J.A.M.A.* **272**, 101–104.
- Cornish BH, Ward LC, Thomas, BJ, Jebb, SA & Elia M (1996): Evaluation of multiple frequency-bioelectrical impedance and Cole-Cole analysis for the assessment of water volumes in healthy humans. *Eur. J. Clin. Nutr.* **50**, 159–154.
- Cox R & Soeters PB (2000): Validation of bio-impedance spectroscopy: effects of degree of obesity and ways of calculating volumes from measured resistance values. *Int. J. Obes. Relat. Metab. Disord.* **24**, 271–280.
- De Lorenzo A, Candeloro N, Andreoli A & Deurenberg P (1995): Determination of intracellular water by multifrequency bioelectrical impedance. *Ann. Nutr. Metab.* **39**, 177–184.
- Detsky AS, Naylor DC, O'Rourke K, McGeer AJ & L'Abbe KA (1992): Incorporating variations in the quality of individual randomised trials into meta-analysis. *J. Clin. Epidemiol.* **45**, 255–265.
- Deurenberg P (1996): Multifrequency-bioelectrical impedance a comparison between the Cole-Cole modelling and Hanai equations with the classical impedance index approach. *Ann. Hum. Biol.* **23**, 31–40.
- Deurenberg P, Westrate I, Paymans I & Koy K (1988): Factors affecting bioelectrical impedance measurements in humans. *Eur. J. Clin. Nutr.* **42**, 1017–1022.
- Ellis KJ (1996): Whole-body counting and neutron activation analysis. In *Human Body Composition*, eds. AF Roche, SB Heymsfield & TG Lohman, pp 45–61. Champaign, IL: Human Kinetics.
- Ellis KJ & Wong WW (1998): Human hydrometry: comparison of multifrequency bioelectrical impedance with ²H₂O and bromide dilution. *J. Appl. Physiol.* **85**, 1056–1062.
- Emerson JD & Colditz GA (1983): Use of statistical analysis in *The New England Journal of Medicine*. *New Engl. J. Med.* **309**, 709–713.
- Evans M & Pollock AV (1985): A score system for evaluating random control clinical trials of prophylaxis of abdominal surgical wound infection. *Br. J. Surg.* **72**, 256–260.
- Forbes GB (1987): Techniques for estimating body composition. In *Human Body Composition*, pp 5–45. New York: Springer.
- Gardner MJ, Altman DG, Jones DR & Machin D (1983): Is the statistical assessment of papers submitted to the *British Medical Journal* effective? *Br. Med. J. Clin. Res.* **286**, 1485–1488.
- Gardner MJ, Machin D & Campbell MJ (1986): Use of checklists in assessing the statistical content of medical studies. *Br. Med. J.* **292**, 810–812.
- Gonzalez C, Evans JA, Smye SW & Holland P (1999): Variables affecting Bioimpedance analysis measurements of body water. *Med. Biol. Engng. Comp.* **37**, 106–107.
- Goodman SN, Berlin J, Fletcher SW & Fletcher RH (1994): Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Ann. Intern. Med.* **121**, 11–21.
- Hannan WJ, Cowen SJ, Plester CE, Fearon KVH & Debeau A (1995): Comparison of bioelectrical impedance spectroscopy and multi-frequency bioimpedance analysis for the assessment of extracellular and total body water in surgical patients. *Clin. Sci.* **89**, 651–658.
- Hardy RJ & Thompson SG (1998): Detecting and describing heterogeneity in meta-analysis. *Stat. Med.* **17**, 841–856.
- Ho TL, Kushner RF, Schoeller DA, Gudivaka R & Spiegel DM (1994): Bioelectrical impedance analysis of total body water in haemodialysis patients. *Kidney Inte.* **46**, 1438–1442.
- Hoffer EC, Meador CK & Simpson DC (1969): Correlation of whole-body impedance with total body water volume. *J. Appl. Physiol.* **27**, 531–534.
- Jebb SA & Elia M (1993): Techniques for the measurement of body composition: a practical guide. *Int. J. Obes. Relat. Metab. Disord.* **17**, 611–621.
- Jenicek M (1989): Meta-analysis in medicine. Where we are and where we want to go? *J. Clin. Epidemiol.* **42**, 35–44.
- Johnson DW, Thomas BJ, Fleming SJ, Westhuyzen J, Moran D & Ward LC (1996): Monitoring of extracellular and total body water during haemodialysis using multifrequency bioelectrical impedance analysis. *Kidney Blood Press Res.* **19**, 94–99.
- Kelly S, Berry E, Proderick P, Harris KM, Cullingworth J, Gathercole L, Hutton J & Smith MA (1997): The identification of bias in studies of the diagnostic performance of imaging modalities. *Br. J. Radiol.* **70**, 1028–1035.
- Khan KS, Daya S & Jadad A (1996): The importance of quality of primary studies in producing unbiased systematic reviews. *Arch. Intern. Med.* **156**, 661–666.
- Kushner RF, Gudivaka R & Schoeller DA (1996): Clinical characteristics influencing bioelectrical impedance analysis measurements. *Am. J. Clin. Nutr.* **64**, 423–427.
- Lionel NDW & Herxheimer A (1970): Assessing reports of therapeutic trials. *Br. Med. J.* **3**, 637–640.
- Lykken GI, Lukaski HC, Bolonchuck WW and Sandstead HH (1983): Potential errors in body composition as estimated by whole body scintillation counting. *J. Lab. Clin. Med.* **4**, 651–658.
- McMaster University Health Sciences Centre (1981): How to read clinical journals: III. To learn the clinical course and prognosis of disease. *Can. Med. Assoc. J.* **124**, 869–872.
- Moher D, Jadad AR, Nichol G, Penman M, Tugwell P & Walsh S (1995): Assessing the quality of randomised controlled trials: an annotated bibliography of scales and checklists. *Control Clin. Trials.* **16**, 62–73.
- Moher D, Jadad AR & Tugwell P (1996): Assessing the quality of randomised controlled trials. *Int. J. Tech. Assess. Health Care.* **12**, 195–208.
- Mulrow CD, Linn WD, Gaul MK & Pugh JA (1989): Assessing quality of a diagnostic test evaluation. *J. Gen. Intern. Med.* **4**, 288–295.
- National Institutes of Health Technology Assessment Conference Statement (1994): Bioelectrical impedance analysis in body composition. *Am. J. Clin. Nutr.* **64**, 524S–532S.

- NHS Centre for Reviews and Dissemination (1996): *Undertaking Systematic Reviews of Research on Effectiveness. CRD Guidelines for those Carrying Out or Commissioning Reviews* (CRD report no. 4) York: NHS Centre for Reviews and Dissemination, University of York.
- Oxman AD (1994): Checklists for review articles. *Br. Med. J.* **309**, 648–651.
- Pullicino E, Coward WA & Elia M (1992): The potential use of dual frequency bioelectrical impedance in predicting the distribution of total body water in health and disease. *Clin. Nutr.* **11**, 69–74.
- Reisch JS, Tyson JE & Mize G (1989): Aid to the evaluation of therapeutic studies. *Paediatrics* **84**, 815–827.
- Rochon PA, Gurwitz JH, Cheung CM, Hayes JA & Chalmers TC (1994): Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal. *J.A.M.A.* **272**, 108–113.
- Schoeller DA (1996a): Bioelectrical impedance analysis for the measurement of human body composition: where do we stand and what is the next step? *Nutrition*, **12**, 760–762.
- Schoeller DA (1996b): Hydrometry. In *Human Body Composition*, eds. AF Roche, SB Heymsfield & TG Lohman, pp 25–44. Champaign, IL: Human Kinetics.
- Schoeller DA, Kushner RF, Taylor P, Dietz WH & Bandini L (1985): Measurement of total body water: isotope dilutions techniques. In *Body Composition Assessment in Youth and Adults: Sixth Ross Conference on Medical Research*, ed. AF Roche, pp 124–129. Columbus, OH: Ross Laboratories.
- Streiner DL & Norman GR (1995): From items to scales. In *Health Measurement Scales: a Practical Guide to their Development and Use*, pp 85–103. Oxford: Oxford University Press.
- Van Marken Lichtenbelt WD, Snel YEM, Brummer RJM & Koppeschaar HPF (1997): Deuterium and bromide dilution and bioelectrical impedance spectrometry independently show that growth hormone-deficient adults have an enlarged extracellular water compartment related to intracellular water. *J. Clin. Endocrinol. Metab.* **82**, 907–911.
- Visser M, Deurenberg P & Staveren WA (1995): Multi-frequency bioelectrical impedance for assessing total body water and extracellular water in elderly subjects. *Eur. J. Clin. Nutr.* **49**, 256–266.
- Westerterp KR, Wouters L & Van Marken Lichtenbelt WD (1995): The Maastricht protocol for the measurement of body composition and energy expenditure with labeled water. *Obes. Res.* **3**, 49–57.
- Working Group on Recommendations for reporting of clinical Trials in the Biomedical Literature (1994): Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. *Ann. Intern. Med.* **121**, 894–895.

Appendix 1

Test to evaluate BIA studies

Study number	Reader number	Date (ddmmyy)		
A. Study design				
1. What type of study design was used?				
Case report				
Uncontrolled experiment				
Cohort or case control study				
Not clear				
2. Were the inclusion criteria clarified?				
		Yes	No	NC
3. Were the exclusion criteria clarified?				
		Yes	No	NC
B. Subjects				
4. Were the recruitment methods of the study group described?				
		Yes	No	NC
5. Was there a description of the demographic characteristics of the subjects? (ethnicity, gender, age)				
		Yes	No	NC
6. Was there a description of the clinical characteristics of the subjects? (normal or type of disease and severity)				
		Yes	No	NC
7. Was the study group free of patients with co-morbid conditions?				
		Yes	No	NC
8. Was the sample selected in a way that avoids bias?				
		Yes	No	NC
C. Procedures				
9. Was the weight measured to the nearest 0.1 kg (1–5 g in children less than 3y)?				
		Yes	No	NC
10. Did the subjects fast overnight or for 8 h before the test? (3 h before in children < 3 y)				
		Yes	No	NC
11. Did the subjects avoid strenuous exercise for 8 h before the test?				
		Yes	No	NC
12. Was the bioelectrical impedance measurements (BIA) performed in a standardised manner? (See questionnaire)				
13. Was the gold standard (GS) performed in a standardised manner? (See questionnaire)				
14. Were BIA and GS tests done at the same time and in similar conditions?				
		Yes	No	NC
15. Was the GS test performed on all subjects?				
		Yes	No	NC
16. Was a test of BIA reproducibility carried out?				
		Yes	No	NC
17. Was non-random exclusion of subjects avoided?				
		Yes	No	NC
18. Was the hydration status likely to have remained the same during the whole measurement period?				
		Yes	No	NC
D. Calculations and statistical analysis				
19. Was the algorithm to calculate the water volumes given?				
		Yes	No	NC
20. Were all results included in the analysis?				
		Yes	No	NC
21. What was the statistical analysis used?				
		Non parametric Correlation/regression Bland/Altman analysis		

NC: not clear.

BIA technique standardisation

Were the methods described in sufficient detail to allow performance of the BIA replication? If there is reference to another study this should be evaluated with the same questionnaire.

Study number	Reader number	Date (ddmmyy)
--------------	---------------	---------------

1. Was the bladder emptied before the test?	Yes	No	NC
2. Was the menstrual cycle considered or not applicable?	Yes	No	NA
3. Was a non-conductive surface used for the test?	Yes	No	NC
4. Was the skin cleaned with alcohol?	Yes	No	NC
5. Was the environment or skin temperature noted?	Yes	No	NC
6. Were metallic objects removed from the subjects?	Yes	No	NC
7. Was the subject position and time spent in that position recorded?	Yes	No	NC
8. Was the height measured to the nearest 0.5 cm? (0.1 cm in children under 3 y)	Yes	No	NC
9. Was the abduction of limbs described?	Yes	No	NC
10. Was the electrode position described? (In children under 3 y the minimal distance between the sense and voltage electrodes should be 3 cm)	Yes	No	NC
11. Were the brand or size of electrodes stated?	Yes	No	NC
12. Did the study state the side of the body on which measurements were taken?	Yes	No	NC

NC: not clear. NA: not applicable.

Dilution technique standardisation

Were the methods described in sufficient detail to allow performance of the GOLD STANDARD replication? If there is reference to another study this should be evaluated with the same questionnaire.

Study number	Reader number	Date (ddmmyy)
--------------	---------------	---------------

1. Were drinks avoided for several hours before the test?	Yes	No	NC
2. Were baseline physiologic samples taken before the dose of tracer?	Yes	No	NC
3. Did each subject take the complete dose? (The dose bottle should be washed out with water and this water should be given to the subject)	Yes	No	NC
4. Was the reported period of equilibration adequate?	Yes	No	NC
TBW: 3–4 h			
ECW: 3–4 h			
If ECW expanded TBW: 4–5 h			
If ECW expanded ECW: 5–6 h			
5. Were the samples stored in airtight tubes?	Yes	No	NC
6. Was the measurement procedure the same for all subjects?	Yes	No	NC
7. Was eating or drinking avoided during the equilibration period?	Yes	No	NC
8. If the back-extrapolation method was used were foods avoided for 1h after the dose or not applicable?	Yes	No	NA
9. If urine was sampled for TBW was the first specimen discarded and then two samples collected after the dose or it was not applicable?	Yes	No	NA

NC: not clear. NA: not applicable.

TBK standardisation

1. Did the subjects change the street clothes before the test?	Yes	No	NC
2. Was jewellery from the subject removed?	Yes	No	NC
3. Was the calibration of the instrument with phantoms confirmed?	Yes	No	NC