PRINCIPLES, PROBLEMS AND A PARADOX WITH THE MEASUREMENT OF ENERGY EXPENDITURE OF FREE-LIVING SUBJECTS USING DOUBLY-LABELLED WATER

J. R. SPEAKMAN
Department of Zoology, University of Aberdeen, Aberdeen, AB9 2TN, Scotland, U.K.

SUMMARY
Many aspects of the study of human nutrition would benefit from the ability to measure the energy which is expended by subjects as they go about their normal activities. The doubly-labelled water technique is a method which allows such measures to be made. The technique depends on the principle that a measure of carbon dioxide production is possible from the difference in the rates at which isotopic labels of hydrogen and oxygen are eliminated from the body. This simple explanation however conceals a host of assumptions. Several issues have emerged as important in the application of the technique to humans. These issues include first, the use of two samples (an initial and final sample) or multiple samples to estimate the rates of isotope elimination, and secondly the appropriate use of dilution spaces. This paper reviews the current status of the debates concerning these two issues. Paradoxically, improvements in our understanding of the technique, in the methods for calculating carbon dioxide production and in the accuracy of mass spectrometry, have not led to improvements in the accuracy of the technique. The mean deviation of estimates of carbon dioxide production using the technique from estimates made by conventional methods averages 7 per cent.

INTRODUCTION
Many outstanding problems in the study of human nutrition revolve around the difficulties of obtaining accurate measurements of the energy which is expended by subjects as they go about their normal behaviour. For example, obesity is a topical nutritional problem which has wide ranging consequences for health. It is obvious that obesity is a problem of energy balance, subjects are either eating too much or expending too little (or both). Debate has raged however as to the relative contribution that each of these causes makes to the development of obesity. The basic problem is that until recently we have not been in a position to test which of the potential causes is of most importance. This is because we did not have a technique which was capable of measuring the energy which is expended by obese people, as they go about their daily lives, compared to the energy expended by lean people.

The traditional methods which have been used to measure energy expenditure are direct calorimetry, and the inference of heat production from oxygen consumption or carbon dioxide production (indirect calorimetry). Both these approaches require confinement of the subject in a relatively small chamber, in order to collect the respiratory gases or to trap the evolved heat. Although these chambers provide useful information on the basal levels of energy requirement, by their very nature they can never fully reproduce the multiplicity of factors which influence daily energy demands.

An alternative to the calorimetry approach is to infer respiratory energy expenditure from the energy which is consumed as food minus any losses as faeces, urea and growth or fat deposition.

0277-6715/90/111365-16$08.00
© 1990 by John Wiley & Sons, Ltd.
This latter method, however, is critically dependent on the ability to accurately weigh, and assess the composition of all the ingested food. When this is left to the subjects themselves they frequently under-record their intakes. More importantly such under-declaring, leading to low inferred energy expenditure, may vary in relation to the critical variable under examination, for example obesity. This may lead to spurious inferences concerning the cause of the problem under study.

The doubly-labelled water (DLW) technique is a non-invasive method which allows an estimate of the energy expended by subjects as they go about their normal activities. In humans the technique provides an estimate of energy expenditure averaged over a period of a few days to a month. In small animals the period is one to two days. The technique has several advantages over the traditional methods of direct and indirect calorimetry. A most important advantage is the ease with which it is performed. It involves only the provision of a daily urine sample, hence subjects can engage in a completely natural daily routine. As such the technique has an enormous potential and is already providing insights into problems of energy balance in the study of human nutrition.

This review is in three parts. In the first part I will review the basic principles and assumptions which underly the technique as it was originally developed for the study of small animals. The application of the technique to humans has stimulated an exponential growth in investigations into the technique. It has also revealed several new problems. In the second part of this paper I will discuss two of these problems. This is not an exhaustive treatment since several other problems exist, however these two problems have had a major role in recent debates about the technique. Finally, I will introduce a paradox which is evident in validations of the technique.

PRINCIPLES

Although the DLW technique is often described by human nutritionists as a new technique, the original work which led to its development was done half a century ago. The technique is only new therefore to the extent that it is only within the last few years that it has been applied to the study of human nutrition.

The experiments which led to development of the technique were performed to establish the fate of inspired oxygen, and the source of oxygen in expired CO₂. It had been suggested that inspired oxygen combined with tissue hydrogen to form water, while the oxygen in carbon dioxide was derived half from the substrate which was being oxidized (fat, protein or carbohydrate) and half from the body water. Lifson et al. recognized that if an animal was forced to breathe oxygen which had been enriched with a heavy isotope of oxygen (¹⁸O), then the isotopic label would enter the body water, and, because the oxygen in expired CO₂ is derived half from the body water, would also appear in the expired CO₂. If the other half of the oxygen in expired CO₂ was derived from the metabolic substrate they suggested that isotopic exchange between oxygen in the substrates of the body and oxygen in body water would bring the enrichment of ¹⁸O in CO₂ close to that in the body water. Furthermore, exchange of oxygen isotopes between CO₂ and water during transport of the gases in the blood would enhance this parity. In an earlier experiment the appearance of inspired oxygen in CO₂ had been inferred to suggest that oxygen is directly involved in substrate oxidation.

Lifson et al. confirmed that when an animal (mouse) is forced to breathe heavy oxygen the ¹⁸O does appear in the expired CO₂, but as they predicted it also appears in the body water at the same enrichment. Moreover, if an animal is injected with water enriched in ¹⁸O, the label also appears in the expired CO₂, meaning the appearance of ¹⁸O derived from inspired air in expired CO₂ could be completely accounted for by equilibration of the ¹⁸O in water.
Lifson's great insight was that the equilibration of respiratory carbon dioxide with body water, that is the fact that respiratory CO$_2$ has the same isotopic enrichment of $^{18}$O as body water, held the potential to measure CO$_2$ production and hence energy expenditure. Lifson saw that the turnover of oxygen isotope in the body would depend predominantly on the fluxes of both water and CO$_2$. However, a label of hydrogen would turn over mainly as a function only of the water flux. The difference in the turnover of the two isotopes multiplied by the size of the body compartment in which the isotopes turned over, would then be a measure of CO$_2$ flux. Generally

\[ r_{\text{CO}_2} = \frac{1}{2} N (k_o - k_d) \]  

(1)

where $k_o$ and $k_d$ are the fractional turnover rates of the oxygen and hydrogen (termed $k_d$ to indicate that label is deuterium and not tritium) per unit time, $N$ is the body compartment size, which was equated with the body water pool, $1/2$ is a constant which accounts for the fact that the turnover of a water molecule reflects the loss of only half a molecule of CO$_2$ since there is only one atom of oxygen in each water molecule, compared with two atoms in each CO$_2$ molecule, and $r_{\text{CO}_2}$ equals the production of CO$_2$ per unit time. Because the decline in the abundance of the isotopic labels in the body is exponential, approaching ever nearer to the natural abundance as time goes on, the fractional turnover rates, $k_o$ and $k_d$, are calculated from the gradient of the decline in the log converted enrichments of the isotopes above the background natural abundance. An estimate of the size of the pool, $N$, in which the isotopes turn over can also be made from measurements of isotopic enrichment in the body. If the amount of isotope used in the dose is known then the extent to which it is diluted by the body is a measure of how much dilutant is present in the body. There are two alternative ways in which this calculation can be performed – using either the initial measured isotopic abundances after a period of equilibration, called the plateau method, or by back extrapolation of the turnover line to the intercept at the initial time when the isotopic dose was administered, called the zero-time intercept technique.

The elegance of the DLW method, which was recognized from the outset, is that the fractional turnover rates may be assessed by injecting the animal with both isotopes, and, after a short period for equilibration, by collecting samples of body water (plasma or urine) at the start and end of a measurement period (Figure 1). The subject could then engage in activities between the taking of the samples unencumbered by the restrictions of calorimetry apparatus. Because two isotopic labels in water are used the technique became known as the DLW technique.

During the 1950s, Lifson and co-workers performed several validation studies of the technique on small laboratory mammals, by comparison with indirect calorimetry, see, for example, Reference 3. In addition they detailed and investigated the various assumptions which underlie the simple model of isotope behaviour represented by equation (1). This work culminated in a major review of theory of the technique.\footnote{Lifson and McClintock identified five major assumptions in the use of the technique (Table 1).}

Lifson and McClintock identified five major assumptions in the use of the technique (Table 1). The first assumption, that $r_{\text{CO}_2}$ is constant over the measurement period, is violated in most animals, since no animal maintains a constant energy expenditure over any protracted period of time. However, it was shown that the simple equation (1) is robust to such variations, and results in a mean estimate of energy expenditure over the inter-sample interval. This is because the route of decay taken by the isotopes, resulting from variations in the rate of CO$_2$ production between the two samples, has absolutely no effect on the calculation (Figure 1) since the only values included in the estimate of $k_o$ and $k_d$ are the initial and final isotopic enrichments. In other words the route taken by the isotopes to reach the final point is irrelevant. An analogy may make this clearer. If two climbers set out from the top of a mountain at the same time along different routes, and they arrived at a base camp at exactly the same time, an observer who knew only the time
Figure 1. Schematic of the changes in isotopic enrichment within a body following introduction of labels of both hydrogen and oxygen. Hypothetical tracks taken by the isotopes during different decay scenarios are shown as dashed lines.

Table I. Assumptions in the Lifson and McClintock 1966 model

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The rate of production of carbon dioxide during the measurement period, ( r_{CO_2} ), is constant</td>
<td></td>
</tr>
<tr>
<td>2. The body water pool size, ( N ), is constant during the measurement period</td>
<td></td>
</tr>
<tr>
<td>3. The isotopes in the body are lost from the body only in the form of water and respiratory carbon dioxide</td>
<td></td>
</tr>
<tr>
<td>4. The isotopes which leave the body in water and carbon dioxide do so at the same enrichment as the isotopes left behind</td>
<td></td>
</tr>
<tr>
<td>5. During the experimental period none of the isotopes which leaves the body re-enters it</td>
<td></td>
</tr>
</tbody>
</table>

they had set out and the time they had arrived would calculate an exactly equal average rate of descent for the two climbers. The intricacies of the two different routes the climbers took would be irrelevant to the calculation.

The assumption of constant body water pool size, \( N \), is also unrealistic. However, it was shown\(^4\) that if the initial and final body masses are used to estimate an average pool size over the course of the measurement, effectively assuming that the change in mass throughout the measurement period is linear, then the resultant estimate also represents the mean energy expenditure over the measurement period. Deviations from linear changes in the pool size introduce negligible errors.

The third assumption that the isotopes only pervade the water pool and are only incorporated into water and \( CO_2 \), is also violated in most animals. This is shown by the fact that the body pool size estimated from the dilution space of the isotopes exceeds the size of the body water pool established from desiccation.\(^5\) This means that when the isotopes pervade the body water pool they also exchange reversibly, but to a relatively small extent, with other compounds in the body. It has been established empirically that this small amount of reversible exchange is greater for
hydrogen than for oxygen. Furthermore there is also some incorporation of the isotopes into other body constituents, notably fat but also protein and carbohydrate, and these are effectively irreversible losses (called sinks) to the system since the turnover of these body compartments is very slow, relative to the duration of a typical DLW experiment. Other sinks to the system include incorporation of isotopes into urea and exchange with the contents of the gut, which are then voided. Lifson and McClintock\textsuperscript{4} pointed out that the errors incurred by such violations of the third assumption may be significant. By using estimates for the extent of reversible exchange, revealed by the different dilution spaces for the two isotopes, they calculated an error of up to 5 per cent in the estimated carbon dioxide production might occur. However, they also noted that the effect of the reversible exchange (dilution spaces) was opposite to that resulting from irreversible exchange (sinks). Consequently, when both effects are considered in concert they would to some extent cancel, producing a summed error considerably less than 5 per cent.

Probably the most significant assumption to which Lifson and McClintock paid attention was the assumption that water and CO\textsubscript{2} lost from the body have equal isotopic abundances to the water left behind. This will be violated because of physical fractionation effects, which accompany all changes of phase, and also the possibility of dynamic biological fractionation of the isotopes. The simple equation (1) was reformulated to account for the effects of physical fractionation

\[ r_{\text{CO}_2} = \frac{1}{2} \{ N (k_o - k_d) - (f_2 - f_3) r_o \} / f_3 \]  

(2)

where \( f_1 \) is the physical fractionation of hydrogen at H\textsubscript{2}O liquid to gas change, \( f_2 \) the physical fractionation of oxygen at H\textsubscript{2}O liquid to gas change, \( f_3 \) the physical fractionation of oxygen at CO\textsubscript{2} gas to H\textsubscript{2}O change, and \( r_o \) the proportion of water loss which is fractionated. At the time Lifson and McClintock wrote their review the physical fractionation factors for phase changes in vivo were unknown, hence an approximation was made by assuming that in vitro equilibrium factors at 25°C would be reasonable approximations \((f_1 = 0\cdot93, f_2 = 0\cdot99 \text{ and } f_3 = 1\cdot04)\). Furthermore, the proportion of water loss which is fractionated was also unknown; to minimize error in assuming a value for this latter term they also assumed that half the total water loss, that is, \( \frac{1}{2} k_d N \), is fractionated. Together this yields

\[ r_{\text{CO}_2} = \frac{1}{2} \{ N (k_o - k_d) / 2\cdot08 - 0\cdot0144* N k_d \} \]  

(3)

The constant* derived by Lifson and McClintock of 0\cdot015 for equation (3) is probably a rounding error of the correct value 0\cdot0144. However, since this value depends on the assumptions about the fractionation factors and proportion of fractionated water loss this discrepancy is unimportant.

Finally, the assumption of isotope re-entry was considered and it was suggested that this might represent a significant source of error when there is a build-up of CO\textsubscript{2} around the animal. However, except in special applications, for example to mammals which live in burrows where expired CO\textsubscript{2} may be trapped around the animal, such a build-up is unlikely and consequently in practice the error is probably unimportant.

An application of the DLW technique to measure free-living energy expenditure of a pigeon had already been published by the time Lifson and McClintock published their review.\textsuperscript{6} During the 1970s and early 1980s the numbers of applications and validations of the technique expanded greatly. These studies involved mammals,\textsuperscript{7} birds,\textsuperscript{8} and reptiles.\textsuperscript{9} Most experimental subjects during this period weighed less than 500 g. The reason for this limit on the size of subjects was the cost of the heavy oxygen. In 1975 it was calculated\textsuperscript{10} that to produce an excess enrichment of \( ^{18} \)O of 500 p.p.m. in the body of a 70 kg human would cost 2500 U.S. dollars. This cost prohibited the application of the technique to large subjects.
Two factors changed this situation. First, the cost of isotopes declined. The current cost (1989) to produce an excess enrichment of 500 p.p.m. in a 70 kg human is approximately 2000 U.S. dollars, about a third the cost it was in 1975 in real terms. Secondly, mass spectrometry instrumentation advances meant that it was no longer necessary to enrich to 500 p.p.m. to obtain a reliable estimate, and studies could be made at initial enrichments of 250 and even 100 p.p.m. excess at half and one-fifth the cost, respectively. In real terms, therefore, it is now possible to use the technique for one-fifth to one-fifteenth the cost that it was in the mid-1970s. The first study on humans was published in 1982. This was followed by an exponential expansion in applications and validations of the technique, not only to humans but to a wide range of animals (see the reviews of non-human applications and all validations up to 1988). In addition, further attention was focused on the assumptions underlying the technique, particularly in the human context.

The renewed interest in the theory and assumptions underlying the technique has involved developments in all the areas which were initially recognized as of importance (Table I). Moreover several practical problems peculiar to the application of the technique to humans have emerged as significant. In the second part of this paper I will outline two problems, one concerned with pool sizes, which has wider implications than the applications to humans and the second, the choice of multiple sampling or collection of only two samples, which is probably peculiar to humans and domestic animals.

PROBLEMS

The use of two pools or a single pool

In the Lifson and McClintock equation (3) the isotopic turnover of each label is referred to the volume of body water. This is because the model assumes that the labels pervade only the body water pool. This assumption is approximately correct, and the volume of distribution of the isotopes closely approximates that of the body water pool. However there is a significant discrepancy for both isotopes.

Coward et al. pointed out that to be mathematically correct the total outflow of products containing any label can only be calculated by multiplying the measured isotopic turnover ($k_d$ and $k_o$) by its own dilution space ($N_d$ and $N_o$). They therefore suggested that

$$ r\text{CO}_2 = (k_o N_o - k_d N_d)/2.08 - 0.0144 k_d N_d \tag{4} $$

should be adopted to calculate CO$_2$ production. This differs from (3) principally in the fact that the turnovers are each related to their own dilution spaces. Since $N_d$ exceeds $N_o$ by about 3 per cent this results in a lower estimate of $r\text{CO}_2$ than the original Lifson and McClintock equation. In their paper (Reference 4, Appendix 2) Lifson and McClintock also considered this model but indicated that the difference between the two calculations was only small (5 per cent). The renewed interest in the alternative model occurred because the actual difference between the two equations is a function of the size of $k_d$ in relation to $k_o$ (Table II). As $k_d$ approaches $k_o$ the difference between the two models becomes very large. Indeed at values of $k_d/k_o$ in excess of 0.971 (1/1.03) the difference results in a positive value for CO$_2$ production when using equation (3) but a negative value when using equation (4). Humans have generally higher ratios of $k_d/k_o$ (Table II) and hence the significance of the alternative models for calculating $r\text{CO}_2$ increased as the technique became applied to humans. It is also apparent that the significance of this problem also depends on the field of study. Burns patients for example have very high ratios of $k_d/k_o$ which
Table II. The effect of the ratio $k_d/k_o$ on the difference in calculated CO$_2$ production when using the single pool and two-pool models. The percentage difference was calculated as the single pool estimate of carbon dioxide production (equation (3)) minus the two-pool estimate (equation (4)), divided by the single pool estimate and multiplied by 100.

<table>
<thead>
<tr>
<th>$k_d/k_o$</th>
<th>Percentage difference</th>
<th>Representative groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>$\infty$</td>
<td>Fish</td>
</tr>
<tr>
<td>0.95</td>
<td>57.0</td>
<td>Burns patients</td>
</tr>
<tr>
<td>0.90</td>
<td>37.0</td>
<td>Arthropods</td>
</tr>
<tr>
<td>0.85</td>
<td>17.0</td>
<td>Infants</td>
</tr>
<tr>
<td>0.80</td>
<td>12.0</td>
<td>Reptiles</td>
</tr>
<tr>
<td>0.75</td>
<td>9.0</td>
<td>Adult humans</td>
</tr>
<tr>
<td>0.70</td>
<td>7.0</td>
<td>Birds</td>
</tr>
<tr>
<td>0.65</td>
<td>5.6</td>
<td>Small mammals</td>
</tr>
<tr>
<td>0.60</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>0.55</td>
<td>3.7</td>
<td></td>
</tr>
</tbody>
</table>

reflects their high water loss (M. Goran, personal communication). The choice of calculation method for these patients therefore has a profound effect on the estimated CO$_2$ production.

Although equation (4) more accurately expresses the difference in the turnover of the two labels it does not necessarily more accurately reflect $r$CO$_2$. This is because the product $k_d N_d$ measures the loss of all products containing hydrogen. By including this term in the equation it was effectively assumed that the only irreversible loss of hydrogen was as water. We know this assumption is incorrect and the effects of such secondary irreversible losses are in the opposite direction to the effect of the reversible exchange.

This interrelationship was formalized by considering a model in which there was both a subsidiary hydrogen pool, and a subsidiary hydrogen sink. In that model the subsidiary sink was assumed to emanate from the subsidiary pool. This latter assumption is probably unrealistic, since the primary candidates for the sink are fat synthesis and exchange to urea, while the primary candidates for exchangeable sites are protein. However, the source of the sink makes no difference to the numerical solution (see Appendix).

Total hydrogen turnover may be considered as the sum of the component fluxes

$$N_d k_d = N_{do} k_{do} + N_{de} k_{de}$$  \hspace{1cm} (5)

where $N_{do}$ and $N_{de}$ are the hydrogen pools contiguous with and exclusive to oxygen, respectively, $k_{do}$ is the turnover due to water and $k_{de}$ is the turnover due to the subsidiary flux. Exchange between $N_{do}$ and $N_{de}$ is assumed to be almost instantaneous relative to the other fluxes, and completely reversible.

In the DLW technique we are interested in subtracting from the turnover of oxygen ($k_o N_o$) the turnover of hydrogen which is due to water flux ($k_{do} N_{do}$). Rearranging (5) and subtracting gives, ignoring fractionation effects

$$rCO_2 = \frac{1}{2} \{ k_o N_o - (N_d k_d - k_{de} N_{de}) \}.$$  \hspace{1cm} (6)
Table III. Irreversible losses of hydrogen (mols/day) from the bodies of four species

<table>
<thead>
<tr>
<th>Species</th>
<th>Bat</th>
<th>Mouse</th>
<th>Fat rat</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Water</strong></td>
<td>0·480</td>
<td>0·800</td>
<td>3·812</td>
<td>365·5</td>
</tr>
<tr>
<td>Irreversible exchange to urea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urea</td>
<td>0·0212*</td>
<td>0·020</td>
<td>?</td>
<td>1·71</td>
</tr>
<tr>
<td>Exchangeable H†</td>
<td>0·0023</td>
<td>0·0025</td>
<td>?</td>
<td>0·47</td>
</tr>
<tr>
<td>Loss of H in urea as percentage of water H (%)</td>
<td>0·48</td>
<td>0·312</td>
<td>?</td>
<td>0·47</td>
</tr>
<tr>
<td>Incorporation into fat during fat synthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fat</td>
<td>?</td>
<td>0·0458</td>
<td>0·51‡</td>
<td></td>
</tr>
<tr>
<td>Loss of H in fat as percentage of water H (%)</td>
<td>?</td>
<td>1·25</td>
<td>1·20</td>
<td>0·14</td>
</tr>
<tr>
<td>Total losses of non-aqueous hydrogen as percentage of aqueous loss of hydrogen (%)</td>
<td>&gt; 0·48</td>
<td>1·562</td>
<td>&gt; 1·20</td>
<td>0·61</td>
</tr>
</tbody>
</table>

* Calculated from urea concentration
† Assuming only 11 per cent of urea H exchangeable
‡ Estimated from Reference 17, assuming turnover during typical 10 day DLW study is 10 per cent of the fat tissue

Since by definition \( N_e = N_{de} \), and empirically \( N_{de} = 0·03N_e \),

\[
rCO_2 = \frac{1}{2}(k_eN_e - (Ndk_d - 0·03k_{de}N_{de})). \tag{7}
\]

Equation (7) more correctly expresses the turnover of oxygen minus the contribution to that turnover which is due to water than either of the equations (3) or (4), which represent the single pool and two-pool models, respectively. However, it contains an unmeasurable term \( k_{de}N_{de} \). We may therefore ask under what conditions the two equations approximate equation (7) by substituting each of them into that equation. This yields the solution that the single pool model measures the true CO₂ production when \( k_d = k_{de} \), and the two-pool model does so when \( k_{de} = 0 \). This latter result is consistent with formulation of the two-pool model which assumes all irreversible losses of hydrogen are as water.

These conditions may be redefined in terms of the total hydrogen flux, \( rH_2 \), the water hydrogen flux, \( rH_2O \), and the subsidiary hydrogen flux, \( rH_2x \). This redefinition suggests the two-pool model will be accurate when \( rH_2x = 0 \) and the single pool model when \( rH_2x = 0·03rH_2O \). The balance of superiority between these two approaches therefore lies at a point when the subsidiary losses of hydrogen are 1·5 per cent of the loss as water, if \(^2\text{H}\) overestimates body water, as estimated by \(^18\text{O}\), by 3 per cent.

There are relatively few measures of the magnitude of the subsidiary losses of hydrogen although Table III summarizes some of the available data. For a much fuller discussion see Reference 17. For humans it would appear subsidiary losses are well below 1·5 per cent of the water loss and hence the two-pool model is indicated. For small mammals, however, the problem remains unresolved and clearly further quantification of such losses is needed. Moreover this
analysis assumes the ratio of estimated dilution spaces is fixed at 1.03. In practice actual measured
dilution spaces average this ratio but individuals differ in the range 1.01 to 1.07. Whether this
represents biological variability or analytical error is currently under debate. However, if it turns
out that the differences reflect biological variability it is potentially the case that different models
will be appropriate for different individuals of the same species.

The use of two samples or multiple samples

When the first application of the technique was made to humans a protocol was employed
which had remained effectively unchanged since Lifson et al.'s first validation study. That is,
isotopes were administered, after an equilibration period a sample of body water was collected,
the subject was then allowed a period of free behaviour after which a second sample of body water
was taken. The fractional turnover of labels was then calculated from the difference in the log
converted initial and final excess enrichments.

Two years later a novel approach was used. In that study the isotopes were administered and
an initial sample taken, as in previous studies, but samples were also collected at regular intervals
throughout the four days between equilibrium and the end of the experiment. Fractional
turnovers were then estimated by fitting a straight line to the sequential evaluations of isotope
enrichment by the usual least squares procedure. Subsequent applications to humans have either
adopted this multiple sampling and curve fitting approach or have used the two-sample
methodology, and a debate has ensued as to which method is superior.

In theory the precision of the estimates of the fractional turnover rate will be greater with the
multiple sampling approach. This is because there is an analytical error associated with each data
point, and an error due to random variations in the background abundance. With the two-point
technique such errors, particularly in the final sample, where the excess enrichment is close to
background, pass directly to the estimate of fractional turnover. With the multiple sampling
technique the estimate of gradient is unbiased by random analytical errors and background
variations.

However, the novel approach also revealed that the decay curve of isotope enrichment is not
a smooth exponential but is highly variable – sometimes sequential isotope enrichments even
increased. Subsequent curves have not involved the same extent of variation, but some variation
is always apparent and it is generally inferred to be biological in nature. Coward and Prentice
stated '... using the two-point method estimates of precision can only be made on the basis of
analytical errors because the effect of biological variations cannot be taken into account. Levels of
imprecision may thus be underestimated.' Our italics added.

The reason the effects of non-random biological variations on the behaviour of the isotopes
cannot be taken into account is because the necessary analyses between the initial and final
sample are not made. However the two-sample technique is robust to variation in the production
of CO₂ over the interval between the two samples, and it is such variation which leads to
isotopic fluctuations in the decay curve (see also Figure 1). The value resulting from the two-point
method represents the mean CO₂ production over the measurement interval. The failure to
incorporate information on the non-random biological variation therefore has no effect on the
estimated precision and it is not, as suggested by Coward and Prentice, underestimated by
taking into account only the analytical error.

The question arises, however, as to the extent to which these non-random biological variations
in the isotope decay curves, resulting from variation in CO₂ production, increase the error with
the multiple sampling technique. This problem was addressed by computer manipulation of
some two-sample data collected in a field study of insectivorous bats. The manipulations
involved producing hypothetical isotope decay curves based upon realistic scenarios for variation in CO₂ production over 24 hours in small mammals. Straight lines were then fitted to these data and the gradients compared to the unbiased two-sample estimate. These data indicated that diurnal variation in CO₂ production might result in errors of up to 30 per cent in the estimates of total CO₂ production when using multiple sampling. This analysis demonstrated that when non-random variation in CO₂ production is large the two-sample technique is superior.

This is a remarkable result because it suggests that by doing less work one actually arrives at a better answer. To clarify why this might be the case let us return to the analogy of two climbers descending a mountain to a base camp. In the first case an observer knows only the heights of the mountain top and the camp, and the times of departure from the summit and arrival at the camp. The observer can calculate the mean rate of decline from these data, and as detailed previously this estimate is not influenced by the intricacies of the different routes taken by the two climbers if they depart and arrive at the same time. The error in the estimate depends only on the random analytical error in the estimated heights of the summit and camp, we shall assume error in the timing is negligible.

Consider now a second observer who plots the routes taken by the two climbers by taking their heights on the mountain at a series of preset intervals. The route taken by climber one is a direct linear path between the summit and the base camp, which is descended at a constant rate. Climber two, however, rapidly comes down a sheer face and then slowly walks back to the camp along a flat path. The observer plots out the graphs of elevation against time for the two climbers and fits straight lines to them using the standard least-squares procedure. The rates of descent are then estimated from the fitted gradients. For the first climber, who took the direct route at a constant rate, this approach works well and the estimated rate of descent is superior to that made by the first observer who only took two measures. This is because any random errors which occur in the estimation of height cancel out in the estimated gradient along the straight line path. However, for the second climber the estimate from the fitted curve is biased, because the route taken is not linear. That is an a priori assumption of the line fitting approach, that the variables were linearly associated results in an error in the estimate of the gradient of descent when the actual path deviates in a non-random fashion from linearity. In this instance the gradient would be overestimated. Where the error from this source exceeds the analytical error in the estimation of heights, which was the only source of error in the two-point estimate, the first observer, who has done less work, will indeed get a better answer for the rate of descent.

Two criticisms may be levelled at these results. First, over the time course of a typical experiment involving humans, variations in CO₂ production are much less than over the diurnal cycle of a small mammal. Secondly, only the error which occurs in the gradient of the fitted line was taken into account. It is well established that random measurement errors lead to covariation between the gradient and intercept of fitted regressions – larger gradients being associated with larger intercepts. This effect occurs because there is a pivot at the locus (x, y) through which a fitted line must pass. If the estimated gradient increases because of random measurement errors then the intercept will also increase, and vice versa. The intercept is often used to estimate the dilution space of the isotopes (N₀ and N_d) using the zero-time intercept method. Larger intercepts lead to smaller dilution spaces. In equation (4) the turnovers (k₀ and k_d), which are estimated from the gradients, are multiplied by the dilution spaces, which are estimated from the intercepts. Covariance between the gradient and intercept might then be expected to lead to cancelling errors in the estimated products. By considering only error in the gradient, the error in final CO₂ production may have been overestimated by Speakman and Racey.21

In a recent paper the consequences of non-random biological variation in the decay curve were further investigated using the computer modelling approach.22 23 In this paper much lower levels
of variation in CO₂ production were considered – at 1·1 times the average for one-third of the period and 0·95 times average for the remainder. This is the sort of variation which might result from a subject in a sedentary job who is periodically active, for example, at weekends. These variations in CO₂ production were so small that the resultant logarithmic-transformed isotope decay curves were almost indistinguishable from linear, which is straighter than any empirical curve yet published. Moreover, in the recent paper²² the effects of the different approaches on both the gradient and intercept were considered (see also Reference 23).

These simulations produced two results. First, covariance of the gradient and intercept is dependent on the measurement errors occurring at random, because it is only in this condition that the locus \((x, y)\) around which the gradient pivots is stable. With non-random biological variation this locus is not always stable and therefore the errors in gradient and intercept do not always cancel, and indeed sometimes they are additive. Secondly, simply because a line appears straight we should not discount the possibility of errors. Decay gradients with \(r^2\) values greater than 0·999 still produced errors in the final CO₂ production estimate of up to 2 per cent.

The conclusion of these analyses is that the two-sample technique is robust to temporal variation in isotope abundance due to non-random biological variation in CO₂ production, but it is prone to analytical errors and random background variation in the final sample. The effects of such errors in the initial sample are relatively unimportant since it is very highly enriched above background. For multiple sampling the converse is true; random analytical errors and background variation are unimportant, but non-random biological variation may result in errors. Since the importance of analytical error and background variation in the two-sample technique depends on the final excess enrichment, the errors in the two-sample technique can be reduced by increasing the dose of label.

I have modelled this tradeoff in errors by considering the error in \(r\text{CO}_2\) for both techniques as changes occur in the non-random biological variability and the initial isotope dosage. Non-random biological variability was modelled by introducing a kink into the theoretical linear decay curve, reflecting an initial period of low energy expenditure followed by a shorter period of high energy expenditure, and was expressed by the linearity of the decay curve: \(r = 1·000\), no variability, \(r = 0·96\) extreme variability. As the curve becomes progressively kinked, the correlation coefficient which expresses only the strength of linear association declines; 0·96 is the correlation coefficient equivalent to a kink reflecting the typical diurnal variation in CO₂ production of a small mammal. For two-sampling, random background variation over a typical 10–14 day experiment was assumed to be 3 p.p.m.,²⁴ and the analytical accuracy of the final sample was assumed to be 2 per cent of the excess enrichment.²⁵ The initial enrichment was varied between 100 and 2000 p.p.m. excess and the fractional turnovers were assumed to be \(k_o = 0·13\) and \(k_d = 0·09\) per day.

A typical tradeoff is shown in Figure 2. As non-random biological variation increased, the error in the multiple sampling approach also increased, but error for the two sample approach was constant. Consequently, at low levels of non-random biological variation the multiple sampling approach was superior, while at high levels of non-random biological variation the two-sampling technique was best. Between these two clear differences was a small area where the two techniques produced similar errors. By plotting many such tradeoffs a picture was built up of how superiority in the two techniques alters with changes in the initial isotope enrichment and the extent of non-random biological variability (Figure 3). Overlaid on this figure are the approximate areas of human nutrition work and small animal investigations.

This two-dimensional error space suggests that as far as error is concerned non-random biological variability in humans is such that multiple sampling will almost always be a superior method. However, in a few circumstances this superiority will only be marginal (the grey area)
Figure 2. The effects of non-random biological variability in CO₂ production on errors of the multiple sampling approach and the two-sampling approach, with initial enrichments of 200 and 1500 p.p.m.

Figure 3. Two-dimensional error space showing the superiority of the two techniques with changes in the non-random biological variability in CO₂ production and the initial isotopic enrichment. In the space above the curve the two-sampling technique is superior, while in the area below the curve the multiple sampling and curve-fitting approach is better. The grey area represented by the curve indicates no clear superiority.

and in yet others the two-sample approach may be better. The choice of technique, especially in the grey area, will almost certainly depend on factors other than error. In laboratories which are set up for automated isotopic analyses each analysis is relatively cheap and the overall costs will be minimized by reducing the dose and using the multiple sampling approach. In contrast, where analyses must be done manually, or for one-off applications, where isotope analyses must be done commercially, the major cost is in the isotopic analysis. It may be cheaper, in these circumstances, to increase the dose and use the two-sample approach. In other words the practical decisions concerning which technique should be used will frequently depend on more than the theoretical arguments concerning errors. In particular cost may be an important consideration. In some situations the lower error with multiple sampling may not be required and in these circumstances the two-sample approach provides a cheaper and more easily performed alternative.
Figure 4. Algebraic error (made positive) for all validations of the technique by comparison to indirect calorimetry over the period 1955 to 1988

PARADOX

Over the history of the development of the DLW technique there have been a number of significant advances. Improvements have occurred along the technical front, in terms of the precision and accuracy which we can now expect in gas isotope ratio mass spectrometry. Furthermore we have advanced in terms of our understanding of the behaviour of isotopes within the body and at the interface of the body and environment.

Although the five major assumptions (Table I) identified by Lifson and McClintock were also basically the same assumptions considered in the reviews by Mullen and Nagy, and indeed in more contemporary reviews, our knowledge about the effects of violating these assumptions are now much clearer, and we have more information available, for example in vivo fractionation factors, on which to base our conclusions and treatment of experimental data. Recently proposed equations for calculating rCO₂ yield significantly different values to the original equation of Lifson and McClintock. The paradox is that despite these technical, empirical and theoretical advances, the mean algebraic error obtained in all the validation studies to date shows no evidence of any decline with time (Figure 4). This could be an artefact of using the algebraic error which assesses only accuracy. A better measure which involves both accuracy and precision is the mean deviation, yet this too shows no evidence of decline over the past forty years (Figure 5). Mean deviation averages about 7 per cent.

The causes of this lack of improvement are unclear. Roberts has indicated the greater ratio of kₐ/kᵦ in humans and its consequences for error (see also Table II). Human validations comprise a major portion of validations since 1982. However, if small animal studies are considered alone there is also no apparent improvement. The implication is that despite all the recent advances we are still missing something. Since improvements have occurred in the technical arena as well as in the constraints of the framework constructed by Lifson and McClintock the most significant future advances may lie outside this theoretical framework which has dominated our thinking about errors in the technique for almost quarter of a century.

Nevertheless we should look on the positive side. A total of more than 30 validation studies have now been performed in a diverse range of subjects and all have indicated that the technique
works with a mean deviation of around 7 per cent. In practical terms with the average number of human subjects limited by cost to around 10 per experimental group this means the technique can separate differences in CO₂ production between homogeneous groups of the order of 8 to 10 per cent. This resolution is sufficient to provide answers to a wealth of previously intractable problems in the fields of clinical and human nutrition. The future lies in reducing this error to bring more and more problems within the techniques domain.

**APPENDIX**

This appendix considers an alternative model in which the subsidiary flux emanates not from the minor pool but from the major pool.

In this situation given

\[ rH_2TOT = rH_2O + rH_2x \]  

\[ N_d k_d = N_d k_{do} + N_d k_{de} \]  

(8)  

(9)

rearranging (9) and subtracting the turnover of hydrogen due to water flux \( N_d k_{do} \) from the turnover of oxygen \( N_o k_o \) gives

\[ rCO_2 = \frac{1}{3} \{ N_o k_o - (N_d k_d - N_d k_{de}) \} \]  

(10)

When \( k_{de} = 0 \) this yields

\[ rCO_2 = \frac{1}{3} (N_o k_o - N_d k_d) \]

which is familiar as the two-pool model equation.

Furthermore, substituting equation (1) into (10) gives

\[ N_o k_o - N_o k_o = N_o k_o - (N_d k_d - N_d k_{de}) \]  

(11)

Since \( N_o = 1.03 N_d \), substitution into (11) gives

\[ 0.03 N_o k_d = 1.03 N_d k_{de} \]  

(12)
Because 
\[ r_{H_2O} = N_d k_d \]
and 
\[ r_{H_2} = N_de k_{de} \]
equation (12) is satisfied if and only if 
\[ r_{H_2} = 0.03 r_{H_2O}. \]
Consequently the conditions in which the one- and two-pool models are satisfied are independent of the source of the subsidiary hydrogen flux.

ACKNOWLEDGEMENTS
I am grateful to the ISCB-10 organizers for their invitation to participate in the mini-symposium on human nutrition. I am also thankful to Sue Roberts and Michael Goran for interesting discussions about the technique, particularly the significance of the \( k_d / k_o \) ratio, and finally to two anonymous referees for their painstaking improvement of the original manuscript.

REFERENCES