The role of technology in the past and future development of the doubly labelled water method

JOHN R. SPEAKMAN*

Aberdeen Centre for Energy Regulation and Obesity, School of Biological Sciences, University of Aberdeen and Division of Appetite and Obesity, Rowett Research Institute, Aberdeen AB24 2TZ, Scotland, UK

(Received 19 October 2004; in final form 20 April 2005)

The doubly labelled water method is an isotope-based technique that is used to measure the energy demands of free-living animals and humans. It is based on the observation that, in the body, the oxygen in carbon dioxide is in complete isotope exchange equilibrium with the oxygen in body water. Hence, a label of isotopic oxygen in body water is eliminated by both respiratory CO₂ and water turnover, whereas a similarly introduced label of deuterium is eliminated only by water flux. The difference in isotope fluxes therefore permits estimation of CO₂ production, which is correlated to energy demands. The doubly labelled water method has been advanced predominantly by technological advances in mass spectrometry. Although it was first described in the 1950s, it was only used on small animals and in low numbers because the costs of the isotopes were a primary constraint. However, advances in mass spectrometry precision and accuracy in the 1980s made it possible to reduce the quantities of isotope used, and hence apply the method on humans, although still in small numbers. The advent of continuous flow inlets in the 1990s made possible the processing of samples in much larger numbers and the sample sizes of studies have expanded. Ironically, however, the technique is now under treat because of technological advances in another area (positron emission tomography), which has generated an enormous demand for ¹⁸O and pushed up the price of isotopes. A continuation of this trend might drive prices to levels where sustained application of the method in human studies is questionable. Replacing determination of isotope enrichments currently performed by isotope ratio mass spectrometry with determinations made by stable isotope infrared laser spectrometry may be a technological advance that will get us out of this problem.

Keywords: Doubly labelled water method; Energy expenditure; Hydrogen-2; IR spectrometry; Isotope analysis; Oxygen-17; Oxygen-18

1. Introduction

The doubly labelled water (DLW) method is an isotope-based technique that measures whole body CO₂ production and hence energy expenditure [1]. The main advantage of the technique over traditional methods of quantifying energy demands is that the method can be employed without the need to restrict subjects inside a calorimetry chamber (either direct to measure
Figure 1. Numbers of publications annually between 1981 and 2004 (January to end of August 2004) on the web of science database that included the phrase ‘doubly labelled water’ or ‘doubly-labelled water’ in the title or abstract. Following a progressive expansion in the 1980s and early 1990s, since 1995, the number of publications has stabilized at around 110 publications each year.

its heat flow or indirect to measure its gas exchange). As such, the method has become the gold standard technique for the measurement of free-living energy demands. Estimates of free-living energy expenditures have applications in many areas, ranging from conservation biology [2] and our understanding of animal responses to climate change [3], to more theoretical ecological studies of limits on animal performance [4, 5]. Within humans, the method is widely employed to quantify the energy demands of various activities or disease states, perhaps none more important than dissecting the aetiology of the obesity epidemic [6]. At present, approximately 120 peer-reviewed manuscripts are published using this technique every year (figure 1). Most of these concern studies of human subjects with relatively few studies of animals.

2. Role of technology in development of the method

The DLW technique has its origins in work performed at the University of Minnesota by Lifson and colleagues in the late 1940s and early 1950s. The primary aim of this work was to solve a long-standing debate over the source of oxygen in respiratory CO₂. Did this come directly from inspired oxygen, or was it, as biochemical studies were increasingly suggesting, derived from water? Clearly, this was a problem that could be addressed by the use of stable isotope tracers. Lifson et al. [7] used enriched gaseous oxygen and labelled water (both provided by the Los Alamos research facility) to trace where the oxygen in expired CO₂ derived from. The original observations were therefore critically dependent on two technological advances. The enrichment procedures for generating enriched compounds that could be used as tracers and the development of the isotope ratio mass spectrometry (IRMS) that could allow isotope levels to be measured in body fluids and gases. Indeed, the importance of the technology in the advent of the method is reflected in the fact that A.O. Neir, the father of modern IRMS, was a co-author of the 1949 Lifson article.

Lifson et al. [7] actually discovered that the oxygen in expired CO₂ was in complete isotope equilibration with the oxygen in body water; hence, any oxygen isotope tracer introduced into
The body water would be eliminated not only by the continuous flux of water through the body, but also by the continual uptake of O\textsubscript{2} and elimination of CO\textsubscript{2}. Lifson recognized that a label of heavy hydrogen would not behave in this manner, because it would only be eliminated by the flux of water. Consequently, if both isotopes were introduced at the same time, their elimination tracks would differ, with the oxygen being eliminated at a faster rate than the hydrogen. The main insight was that this differential elimination holds the potential to measure the rate of CO\textsubscript{2} production. Because the two labels are administered into the body water, the method has become known as the DLW technique.

The first validation of the DLW technique [8] involved 12 mice (\textit{Mus domesticus}). They showed that by making some approximations for the isotope behaviour (notably likely levels of evaporative water losses and fractionation) an estimate of CO\textsubscript{2} production could be generated with an error (accuracy) of \(\sim 4\%\) and a precision of \(\sim 10\%\). The key advance of the technique, however, was that the elimination of the isotopes could be measured from just two samples—taken at the start and end of an appropriate measurement period. Thus, the subjects could roam freely between the samples of the collection, obviating the need for restriction in a conventional calorimetry chamber. It was, however, almost a decade before this potential was realized in a study involving measurement of the energy demands of flight in homing pigeons (\textit{Columba livia}) [9].

Between 1955 and 1975, fewer than 10 studies utilized the DLW method, and the sample sizes of animals used in these studies were generally less than 10 individuals. Why was such an obviously valuable method so under-utilized? The first reason was that the method was expensive. Deuterium and tritium were readily and cheaply available, whereas the \(^{18}\text{O}\) isotope was very expensive. \(^{18}\text{O}\)-enriched water is generated by fractionation. However, because the extent of mass difference of \(^{18}\text{O}\) water from \(^{16}\text{O}\) water is small, the fractionation factor is also small, and therefore many passes in a fractionation column are needed to obtain a sufficient enrichment to use in labelling studies. A second, and related, problem was that the technology for mass spectrometry was extremely specialized and complex. Very few laboratories were capable of performing the requisite analysis, and the errors in isotope enrichment determinations meant that animals had to be labelled with large amounts of isotope to make effective measurements, thereby increasing the cost of the method. Hence, a technique that was enabled by technological advances was effectively limited by the pace at which the mass spectrometry technology developed. This is because the pace of development defined the cost of the studies and, in turn, the ability of researchers to afford it and find facilities capable of doing the IRMS measures. In 1975, it was estimated that to apply the method to a human would cost between 5,000 and 10,000 US$ (costs NOT adjusted for inflation to modern time) [10].

Despite this estimate of prohibitive costs, the first human validation study was published only 7 years later [11] and involved four subjects. What happened during the intervening years was a progressive development in mass spectrometry technology to improve the precision of the machines. Consequently, although Lifson had based his cost estimates on an initial enrichment for \(^{18}\text{O}\) of 2500 ppm (\(\delta^{18}\text{O} \approx 1200\%\)) above background (the enrichment required in studies of animals at that time), the actual enrichment used in the 1982 study was only 250 ppm. Technological advances in the precision and accuracy of mass spectrometry had progressively reduced the required dose; until by the early 1980s, it had become economically feasible to apply the method to the biomedical problems of the day. Although at 600 US$ per subject for \(^{18}\text{O}\) alone, this was still an expensive method to use.

What followed, however, was not an explosion of applications but a gradual expansion in its use over the entire period of the 1980s (figure 1). By 1989, the number of articles that were using the method had expanded to about 20 per annum. Again, technology was simultaneously enabling and holding back the widespread use of the method, because, although the costs of the technique had declined, the preparation of the samples once collected was
an arduous manual task. Furthermore, IRMS was still an extremely specialized technique requiring careful collection of reference data from dual inlet machines, often by manual switchover from reference to sample, and manual collection of the derived beam currents. Once collected, these data had to be manually processed to calculate the derived isotope ratios relative to laboratory working standards. On a good day, a skilled technician could prepare and run 10 samples, but it would probably take another day to perform and check the calculations using an electronic calculator. An additional aspect influencing slow uptake of the technique was that IRMS devices were originally designed with the measurement of natural abundance enrichments of isotopes in the context of earth science studies in mind. The analysis of artificially enriched samples at levels well above the natural abundance range introduced a series of problems that were not widely appreciated until some time later (e.g. the problem of cross-contamination between sample and reference samples in cross-over valves [12]).

The expansion in the use of the method really took off in the early 1990s, until by 1995, about 110 articles per year were being produced. This level has been sustained ever since (figure 1). Several technological advances contributed to this expansion. First, computerization of mass spectrometry apparatus allowed automated sample collection and computer-controlled automated inlets to which multiple samples could be connected. This meant that technical staff could simultaneously prepare samples while the mass spectrometer collected the data and processed it automatically. The second biggest development was the advent of continuous flow analysis, which permitted even greater numbers of samples to be processed. For the stable isotope analysis of oxygen, this development occurred in the early 1990s, but the technological problems of using continuous flow for automation of hydrogen analysis were considerably more significant [13]. In particular, the carrier gas stream of helium (mass 4) was orders of magnitude greater than the sample flow of deuterated hydrogen molecules of mass 3, and interference between these peaks was difficult to control. Attempted solutions to this problem included replacing the carrier gas with argon and having filters that trapped out the helium but allowed the hydrogen to pass, but these were not completely successful [14]. Consequently, there was a large imbalance between the numbers of samples that could be processed for oxygen and hydrogen, which consequently became the rate-limiting step for any application of the method. Nevertheless having only one isotope to worry about in terms of sample-processing time was a great step forward. The primary consequence of this was that the number of subjects in studies started to increase; most studies prior to 1985 had less than 10 subjects, and most studies after 1995 had between 10 and 100.

The solution to the problems of using helium as the carrier gas for hydrogen analysis, and the advent of online pyrolysis machines that performed reduction of water samples online [15], brought processing times for hydrogen samples down to the same time required for oxygen analysis. A second wave of expansion in the technique, however, did not follow this technological advance, although sample sizes in studies continued to creep upwards. The primary reason that there was not a renewed explosion of DLW studies in the early 2000s revolved around a different problem—a crisis in the supply of the $^{18}$O isotope. Initially, this was due to a breakdown in one of the few production facilities, which restricted the availability of the label. By 2000, the cost of $^{18}$O had increased to around 150 US$ per gram $\text{H}_2^{18}$O from a price of less than 70 US$ per gram in 1994. The reduction in supply was, however, combined with a more serious problem, the expansion of a second use for the product. The biomedical uses of $^{18}$O-labelled water are not restricted to its application in the DLW method. In particular, $^{18}$O water is used as a substrate to synthesize $^{18}$F-labelled glucose, which is an essential compound for positron emission tomography (PET) scanning. Prior to the turn of the millennium, PET scanning was at a similar technological situation that mass spectrometry had been in 50 years earlier. It was a specialized method that was expensive and which few people could perform. The use of PET scanning, however, has recently exploded, and this has
sustained the price of $^{18}$O at the same level of 150–170 US$ per gram despite the renewed expansion in production facilities.

The reasons for the explosion in PET scanning include technological developments in the field and the expansion in availability of affordable machines. However, the main factor resulting in the expansion in the number of scans being performed was a change in the policy of US medical insurance companies which early in 2001 made PET scans a reclaimable expense on medical insurance policies. This removed any barrier in terms of cost on people being given diagnostic scans.

3. Future

The problems posed by the technological advance of PET scanning for the DLW method should not be underestimated. The main problem is that the contribution of the cost of $^{18}$O to the total costs of a PET scan is relatively trivial. On average, a 900 US$ scan uses only 15 US$ of isotope. Suppliers of $^{18}$O have not been slow to recognize the potential goldmine that this market represents for their product. There has even been talk by some company representatives (although not all) that the PET market could probably easily stand a price hike in $^{18}$O water from its current level of 150–170 US$ per gram to 450–500 US$ per gram or even higher. For DLW studies of small animals, this would not be a serious issue, because the costs of the isotope analysis remain the highest costs of a measurement. However, for human studies, which comprise probably 90% of current applications, it seems very likely that a 3-fold increase in the cost of isotopes would seriously reduce the number of applications of the method because the costs of obtaining the answers would not justify the gain that the information would provide. The immediate effects might then be a retreat in sample sizes followed by a decline in the number of studies.

4. Solutions to the future problems for the DLW method

There would appear to be several potential solutions that might emerge to save the day for DLW. Two of these include future technological developments. The most obvious immediate solution would be for the production companies to increase supply. However, this is not easily achieved as the fractionation columns producing the enriched isotope have a limited capacity. The only way in which the major producers could increase capacity is to construct new columns or modify existing ones, and either procedure is extremely expensive, costing millions of dollars. Moreover, the advent of a new facility for production at Cambridge Isotope Laboratories in 2003 had virtually no impact on global isotope prices, although by mid 2005, the impact of this production facility had caused a reduction in prices. Given the time before new production plants break even, it seems very improbable that completely new production facilities for $^{18}$O water will enter the market. However, a production facility run by the US government that was mothballed in the early 1990s has recently (2004) been brought back on line as a commercial venture by a company called Spectra Isotopes. Nevertheless, the motivation for doing this is to sell isotopes to PET scanning users who can afford to pay 500 US$ per gram, rather than DLW users who would prefer to pay 100 US$ per gram.

There are perhaps two ways to solve this problem. One is for governments to subsidize the price of the label for DLW studies, or build their own production facilities. The political and commercial problems that government intervention in the isotope-free market economy might lead to are, however, so complex that this may sabotage any development along these lines.
The second solution may be more feasible – that is to move isotope production away from a small group of major producers operating multi-million dollar fractionation columns to a more distributed technology. In fact, gas centrifugation technology may provide a potential route to achieve this. In the future, rather than purchasing $^{18}O$ from a major supplier, we may have bench-top production facilities that could supply the needs of pet scanning and DLW users alike. Whether this is both possible and economically feasible will depend on future advances in gas centrifugation.

When Lifson et al. [10] estimated it would cost 5,000–10,000 US$ to apply the DLW method to humans, the solution to this economic constraint proved to be developments in the precision of mass spectrometry. These developments reduced the dose of isotope required, effectively reducing the costs by an order of magnitude. Could technological advances in the measurement of isotope enrichments deliver a similar solution to the expanding cost of the label? For example, if we could increase the precision of mass spectrometry 10-fold, could this in theory reduce the dose of isotopes needed by another order of magnitude? This might mean that the projected prices of 450–500 US$ per gram of isotope would not be prohibitive on applications of the DLW technique.

Unfortunately, increasing the precision and accuracy of mass spectrometry is unlikely to solve the problem. This is because the precision and accuracy of isotope determinations are probably not limited by mass spectrometry itself, but by the necessary sample processing that is required between sample collection and introduction of the derived gases from water into the analyser. Even if the errors introduced during sample preparation could be reduced, there is another problem that the elevated precision and accuracy cannot overcome. This limiting factor is the fact that there is a natural abundance of the isotopic labels [1, 16]. When applying the DLW method, one measures this background abundance prior to labelling the subjects. Turnovers of the label are then calculated from the excess isotope enrichments above this background. The assumption is made that the background enrichment is constant over the duration of the measurement. However, background enrichments are not constant, so this introduces an error [17], the magnitude of which depends on the isotope dose [16, 17]. Following dosing, one needs to leave the subjects for about two to three biological elimination half-lives to obtain sufficient divergence of the oxygen and hydrogen labels. Consequently, by the time the final sample is taken, it has reduced to $\sim$15–20% of the original dilution of the dose in the body water. Imagine, therefore, that a dose has been administered, which increases the enrichment in body water to 150 ppm above background. By the end of the measurement, the enrichment above background will be $\sim$25–30 ppm. If the background has changed by 0.5 ppm, this would introduce an error in the final isotope determinations of $\sim$1.5–2%. Because the oxygen and deuterium isotopes change in parallel [18], and one is looking at isotope divergence, the actual error in the CO$_2$ measurement introduced by this drift will be lower [16, 17], at around 0.5–1% [19]. However, consider what happens if we were to reduce the initial dose by a factor of 10. In this situation, the initial enrichment would be 15 ppm, and the final enrichment 2.5–3 ppm. A change of 0.5 ppm in the background is now much more serious, comprising an error of 15–20% and an impact on the final calculation of CO$_2$ production and hence energy expenditure of 5–10%.

To escape the problem of potentially increased isotope prices, we therefore need two technological advances, one of which at first sight seems impossible to deliver. First, we not only need increases in the precision and accuracy of isotope determinations, but also need to be able to know how the background isotope enrichments are changing in a subject who is dosed with the label. However, this second problem is perhaps not as intractable as it first appears. We have already mentioned that the background isotope enrichment of $^{18}O$ is paralleled by a change in the deuterium enrichment. This correlation in background enrichment occurs...
because the processes that generate changes in background isotope levels involve fractionation that is common to both isotopes. If we were only interested in background levels of $^{18}$O during $^{18}$O dosing, we could reconstruct these from changes in background deuterium levels. Unfortunately, of course, we have labelled with deuterium as well – but we can use changes in background levels of $^{17}$O to predict the levels of background $^{18}$O because the two correlate almost perfectly [20]. Combining the known relationship of $^{17}$O to $^{18}$O with the relationship of deuterium background to $^{18}$O background [16] gives us the seemingly impossible capability to track how the background is changing while the labels are being eliminated.

There are two practical problems that arise with this idea. Enriching $^{18}$O also enriches $^{17}$O slightly, so isotopes would need to be mixed carefully with a $^{17}$O-depleted product to make sure no $^{17}$O was incorporated as an additional dose. This might elevate the cost of the dose further than any cost benefits that could arise from reducing the total dose. However, the second problem is that measuring $^{17}$O enrichments by mass spectrometry is problematic because the desired gas for making enrichment estimates of $^{18}$O by IRMS is CO$_2$. CO$_2$ has several advantages over water as the analysis compound. Water easily condenses and is difficult to expunge from measurement systems leading to large memory problems. However, in CO$_2$, the mass peak 45 which corresponds to $^{12}$C, $^{17}$O, and $^{16}$O coincides with the mass peak for the isotopologue $^{13}$C, $^{16}$O, and $^{16}$O. Thus, measurements of $^{17}$O enrichment are confounded with measurements of $^{13}$C. Indeed, when measures of $^{13}$C enrichment are made by mass spectrometry, the contribution of $^{17}$O to the measure is eliminated by predicting its enrichment from the enrichment of $^{18}$O. The main difficulty is that the background enrichment of $^{17}$O is $\sim 380$ ppm, but the background enrichment of $^{13}$C is 11,000 ppm. Equivalent percentage fluctuations in $^{13}$C enrichments, therefore, completely swamp changes in $^{17}$O.

The method also assumes that the relationship between background oxygen and hydrogen abundances and the association of $^{17}$O and $^{18}$O conform to the meteoric waters lines and are thus predictable. Such covariance was detected in only four of eight subjects studied by Horvitz and Schoeller [16] but separated from the MWL, and no associations were found by Ritz et al. [19]. Clearly, considerably more information is needed on the patterns of background variation in all three isotopes before the potential of this method might be realized.

5. Is stable isotope ratio infrared spectrometry the answer to these problems?

Stable isotope ratio infrared spectrometry (SIRIS) appears to provide a potential route for solving both of these issues. First, because the measurement depends on the absorption due to rotational vibrations of the molecules and not their masses, the measure of $^{17}$O is completely separated from measurement of $^{13}$C [21, 22]. Indeed, because the measurement medium is water, there is no CO$_2$ present to cause confusion anyway. Secondly, the measurement is made directly on physiological fluids introduced directly into the measuring cell. This potentially avoids any problems of fractionation during offline preparation of the sample, but probably will generate as yet unappreciated problems. If SIRIS units were commercially available, these might provide a route to escaping the threat from the PET scanning community. Moreover, even if the threat of increased prices was not to materialize, adopting the SIRIS method instead of IRMS might improve the precision of the method.

To date, a single study has been performed comparing the performance of SIRIS and IRMS for measuring isotope samples derived from a DLW study [23]. The study species was the Japanese quail (Coturnix japonica), and measurements were made using simultaneous gas exchange and DLW over a period of 24 h. The samples collected for DLW were analysed by the traditional offline preparation and IRMS, as well as by SIRIS. On average, the accuracy of
both methods was good when compared with direct measures of gas exchange – DLW using IRMS overestimated the actual CO₂ production by 1.95%, whereas DLW with the samples analysed by SIRIS averaged a 2.45% overestimate. Individual discrepancies were, however, much greater than these averages, as is the case in all DLW validations published to date. The standard deviation of the differences was 6.5% for both methods.

This low precision of individual DLW measurements might occur because individuals may vary in how closely they approximate the necessary assumptions that are made when deriving the calculation for the DLW method. Hence, these assumptions may be appropriate for the average individual, but for any given individual, there is a discrepancy. For example, one assumption of the technique is that evaporative (and therefore fractionated) water loss is 20% of the total water loss. This may pertain on average across individuals, but any given subject may deviate from this average, leading to discrepancies in the comparison of gas exchange to the DLW calculation. If effects of this type contributed to the precision error in the methods, then we would expect that errors with IRMS would correlate with the errors detected by SIRIS, as they would have a common source. However, this is not the case, as a plot of errors using both methods reveals no significant association (figure 2) \( r^2 = 0.08, p = 0.23 \).

This lack of association in figure 2 suggests that the precision errors in the DLW technique derive not from fundamental errors in the formulation of the model under which DLW estimates are calculated, but rather from analytical errors in the determinations of the isotope abundances used to reconstruct the CO₂ production rates. It would appear from this single comparison that the advantages of using SIRIS (in that the preparation stage is omitted), are offset by the lower precision of the actual determination method. The calculations, however, do not include any attempt to correct for background drift using the \(^{17}\)O determinations [23]. It would be instructive to know if such reconstructions are possible, and what their impact is on the precision of the SIRIS method. In this particular application, which includes a small animal measured over a short time period at relatively high isotope enrichments, the reconstruction of background drift may not be very important. However, in other applications/validations of the method, it may be more significant, so the absence of an effect in this particular study would not necessarily refute the notion of this approach. Moreover, this comparison was made using the SIRIS technology as it stood 5 years ago in 1996. Improvements in the precision of SIRIS may now make it more precise, even if the utilization of the \(^{17}\)O signal proves not to be a useful refinement of the method. Consequently, if the continued increase in prices of \(^{18}\)O

![Figure 2](image-url)  
**Figure 2.** Individual discrepancies between DLW by SIRIS and respirometry estimates of CO₂ production across 18 Japanese quail (\(C. \text{japonica}\)), plotted against the discrepancies between DLW by IRMS and respirometry for the same individuals. Data from van Trigt et al. (2002). The relationship was not significant.
water occasioned by the expansion of PET scanning does not occur in the manner projected, SIRIS may be a more desirable method simply because it reduces the imprecision of individual measurements, if indeed the precision has improved significantly over time. This is particularly the case in applications of the technique where individual rather than group average estimates of metabolic rate assume significance. The machines may also be cheaper than IRMS machines once they become commercially available. Further validations comparing IRMS and SIRIS approaches, particularly in human subjects, are certainly worth pursuing.

Acknowledgements

I am grateful to Erik Kerstel and Livio Gianfrani for the invitation to participate in the SIRIS workshop, and Paula Redman for her very helpful comments on the manuscript.

References
