ESSAY REVIEW

Measuring metabolic rate in the field: the pros and cons of the doubly labelled water and heart rate methods

P. J. BUTLER, *† J. A. GREEN, * I. L. BOYD ‡ and J. R. SPEAKMAN §

*School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK, ‡ Sea Mammal Research Unit, University of St Andrews, St Andrews, Fife KY16 8LB, UK, and § Aberdeen Centre for Energy Regulation and Obesity (ACERO), School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK, and also at ACERO, Division of Energy Balance and Obesity, Rowett Research Institute, Aberdeen AB21 9SB, UK

Summary

1. Measuring the metabolic rate of animals in the field (FMR) is central to the work of ecologists in many disciplines. In this article we discuss the pros and cons of the two most commonly used methods for measuring FMR.

2. Both methods are constantly under development, but at the present time can only accurately be used to estimate the mean rate of energy expenditure of groups of animals. The doubly labelled water method (DLW) uses stable isotopes of hydrogen and oxygen to trace the flow of water and carbon dioxide through the body over time. From these data, it is possible to derive a single estimate of the rate of oxygen consumption ($\dot{V}_\text{O}_2$) for the duration of the experiment. The duration of the experiment will depend on the rate of flow of isotopes of oxygen and hydrogen through the body, which in turn depends on the animal’s size, ranging from 24 h for small vertebrates to up to 28 days in Humans.

3. This technique has been used widely, partly as a result of its relative simplicity and potential low cost, though there is some uncertainty over the determination of the standard error of the estimate of mean $\dot{V}_\text{O}_2$.

4. The heart rate ($f_H$) method depends on the physiological relationship between heart rate and $\dot{V}_\text{O}_2$.

5. If these two quantities are calibrated against each other under controlled conditions, $f_H$ can then be measured in free-ranging animals and used to estimate $\dot{V}_\text{O}_2$.

6. The latest generation of small implantable data loggers means that it is possible to measure $f_H$ for over a year on a very fine temporal scale, though the current size of the data loggers limits the size of experimental animals to around 1 kg. However, externally mounted radio-transmitters are now sufficiently small to be used with animals of less than 40 g body mass. This technique is gaining in popularity owing to its high accuracy and versatility, though the logistic constraint of performing calibrations can make its use a relatively extended process.

Key-words: accuracy, field metabolic rate, precision

Introduction

Ecophysiological and functional ecologists seek to further our understanding of the manner in which organisms operate in their natural environment. Everything that organisms do, both physiologically and behaviourally, involves utilization of energy. Thus, the evolution of organisms is constrained within the envelope of what is energetically feasible. Energy is therefore one of the most important currencies determining the genetic fitness of organisms. All other things being equal, those individuals that obtain and process energy with greatest efficiency, and that balance this against other factors impinging on their survival and reproduction, will have greatest genetic fitness (Tolkamp et al. 2002). The measurement of energy turnover and, in particular, how it is allocated to specific activities, is therefore of central importance to the understanding of the physiological, behavioural and evolutionary ecology of organisms (McNamara & Houston 1996).

© 2004 British Ecological Society

† Author to whom correspondence should be addressed. E-mail: p.j.butler@bham.ac.uk
A wide variety of approaches has been used to measure the energetic state of individual animals (Speakman 1997). In terms of those that are useful in the field, these include the use of time–energy budgets and several different types of field-based ‘respirometry’. In the present paper, we are concerned with comparing the two types of field-based ‘respirometry’ that have been used most widely: the doubly labelled water (DLW) and heart rate ($f_H$) methods. It would probably be fair to say that neither has provided physiological and behavioural ecologists with the tools they really require to investigate questions about the dynamics of animal states. Both include uncertainties and, while there have been many attempts to provide validation of the methods, these validations almost invariably suffer from the problem that they are carried out under very specific conditions on a narrow range of species for a narrow range of their normal activities. Moreover, these methods are indirect measures of gas exchange and the $f_H$ method, at least, needs to be calibrated against more direct measures. Nevertheless, there is no perfect system of measurement and there is a need to continue to improve our understanding of what these respective methods can say about the metabolic rate of animals in the field (FMR). Our aim in this article is to explain how the DLW and $f_H$ methods work, the basic assumptions upon which they rest and the advantages and disadvantages that attend their use.

**UNITS OF MEASUREMENT**

Metabolic rate (MR) is the rate at which energy is expended. However, both the DLW and $f_H$ methods provide estimates of an aspect of gas exchange. Heart rate is normally calibrated against rate of oxygen consumption ($V_O$) and therefore the method gives an estimate of $V_O$ in the field, whereas the DLW method gives an estimate of the rate of CO$_2$ production ($V_{CO_2}$). The volume of oxygen consumed (or CO$_2$ produced) should either be corrected to standard temperature (273 K) and pressure (101·3 kPa or 760 mmHg) dry (STPD) or converted to a mol. One mol O$_2$ occupies 22·39 l at STPD (Dejours 1981). The SI unit for the rate of energy expenditure ($J$ s$^{-1}$) is the watt (W) and the conversion of $V_O$ or $V_{CO_2}$ to W depends on the metabolic substrate. This requires a knowledge of the respiratory quotient (RQ), which is $V_{CO_2}/V_O$ at the cellular level and, in steady-state conditions, will be equivalent to $V_{CO_2}/V_O$ as measured between the animal and its environment. The latter is known as the respiratory exchange ratio (RER) (or RQ, under steady-state conditions). For metabolism of pure carbohydrate (i.e. with a RQ of 1), 1 ml O$_2$ s$^{-1}$ (or 1 ml CO$_2$ s$^{-1}$) = 21·1 W, whereas for pure fat metabolism (i.e. with a RQ of 0·71), 1 ml O$_2$ s$^{-1}$ = 19·6 W and 1 ml CO$_2$ s$^{-1}$ = 27·6 W (Lusk 1919; Brobeck & Dubois 1980). If RQ is unknown, it is often taken to be 0·8 (Schmidt-Nielsen 1997). Thus, if $V_O$ is used to estimate MR, this assumption could lead to an error of +2·5 to −5%, whereas if $V_{CO_2}$ is used, the error could be +9 to −18%.

**Doubly labelled water method**

**PRINCIPLES**

The doubly labelled water method for estimating total CO$_2$ production over a period of time is based on the observation that the oxygen in respiratory carbon dioxide is in isotopic exchange equilibrium with the oxygen in body water (Lifson et al. 1949; Lifson, Gordon & McClintock 1955). This isotope exchange is catalysed by carbonic anhydrase. The speed of the reaction is such that within a few seconds of being produced, any carbon dioxide has the same oxygen isotopic composition as the oxygen in body water. If an isotopic label of oxygen (such as oxygen-17 or oxygen-18) is introduced into body water (either by injection or by drinking a dose of isotopically enriched water) that dose of isotope will be gradually eliminated from the body over time. If the enrichment after dosing is plotted as a function of time, the resultant curve follows a negative exponential. This is because materials such as urine, evaporation and eliminated respiratory carbon dioxide continuously carry the isotopic dose away from the body, while respiratory water, preformed water in food, drinking water and respiratory oxygen, continuously replenish the system with unlabelled water. The rate at which all these materials flow through the body determines how fast the isotope is washed out, but this relationship also depends on the size of the body water pool. Larger body water pools flush out more slowly than smaller pools, everything else being equal. In fact if the logged isotope enrichment above background is plotted against time, the gradient of this linear relationship (called $k$) multiplied by the body water pool size gives a quantitative estimate of the total flow of materials through the body that carry the isotopic oxygen signal. If we know the actual amount of isotope introduced into the body, then the body water pool size (N$_O$) can be estimated from the same curve by the dilution principle. There are two ways to achieve this – first by back extrapolating the elimination curve to the intercept at the injection/dosing time ‘intercept method’, or using a single observed post-dose isotope enrichment ‘plateau method’ (see Fig. 1).

Dosing a Human or other animal with labelled oxygen and tracking the decline in subsequent isotope enrichment can therefore tell us the total flux of oxygen through the body, but being a combination of the CO$_2$ and water flux this is a fairly useless measure. However, because water contains hydrogen, and carbon dioxide does not, introducing a label of hydrogen (such as deuterium or tritium) into the body water and tracking its elimination would enable a measure of the water flux alone. Consequently if one were to introduce both isotopes at the same time (hence doubly labelled water) and measure the elimination of both isotopes, an estimate of CO$_2$ production could be determined by difference (Fig. 1).

In practice, the method involves two different types of protocol that generally depend on the size of the
subject and their amenability to manipulation and sampling. In its simplest form the method requires only that the subject be dosed in a quantitative manner (generally by injection in animals and orally in Humans) and that a sample of body water (generally blood in animals and urine in Humans) is taken once this dose has reached complete isotopic distribution in the body pool – the time for this to occur depends on body size and is now reasonably well quantified (Speakman 1997), being around 30 min to 1 h for animals weighing less than 100 g, around 3–4 h for animals weighing 30 kg and around 6 h for Humans and animals weighing 70–100 kg. In theory it might be anticipated that route of administration would impact on the equilibration time, with intravenous being faster than intraperitoneal or intramuscular and the slowest being oral dosing. Theoretical mixing models (Speakman et al. 2001b), however, suggest this hierarchy may only pertain to penetration of the dose into the plasma pool. From there the isotopes must penetrate the extra- and intracellular water and this equilibration is largely independent of the route of isotope administration. As long as the penetration to the plasma pool is more rapid than mixing into the intra- and extra-cellular water pools, the route of administration will be largely irrelevant to the isotope equilibration time.

Once the isotopes reach equilibrium and are sampled, the subject is allowed to go about its routine business for a variable time period (again depending on body mass this can vary between 24 h for the smallest vertebrates, to 28 days in Humans). After recapture, the subject is sampled again to obtain a final isotope enrichment measure. Since the isotope enrichments need to be expressed relative to background enrichment, some estimate of background isotope enrichment is also needed. Speakman & Racey (1987) defined four different ways to achieve this, but in practice for large animals this involves taking a predose sample of body water and for smaller animals involves sampling some unlabelled individuals that are not part of the experimental measures. This simple method has been termed the ‘two-sample’ methodology (Schoeller et al. 1986), although for larger animals it clearly involves three samples because an individual background is also taken.

The second protocol involves making repeated measures of isotope enrichment during the elimination period and estimating the elimination gradients from fitted curves. This method has been called the ‘multiple-sampling curve fitting’ approach. Since this involves repeated disturbance of the subjects, the only applications of this method have been in studies of Human nutrition (e.g. Prentice et al. 1985). Some hybrid methods are also in use (Westerterp, Wouters & van Marken Lichtenbelt 1995; Delany et al. 1995) again for Human studies and have not yet been applied to animals. Given the requirement for multiple recaptures, it seems unlikely these protocols will spread to ecological investigations of even large animals.

ASSUMPTIONS

The simplified description of the principles of the DLW technique (above) conceals a host of assumptions that were partially elaborated by Lifson et al. (1955) and more fully explored by Lifson & McClintock (1966). Subsequent summaries of the method (Mullen 1973; Nagy 1980, 1983; Speakman 1997) have explored these assumptions in increasingly greater detail. As might be expected, it has become apparent that some assumptions represent greater problems than others, while modifications of the calculation method have enabled the errors associated with some of the assumptions to be eliminated. As our understanding of the method has evolved, some issues that were not even discussed by Lifson & McClintock (1966) have moved centre stage in terms of their importance for the accuracy and precision of the method.

The first assumptions are that all the flows of materials and the sizes of the respective pools are constant throughout the measurement period. This is clearly not going to be true for any living system that we might be interested in measuring. Speakman & Racey (1986) showed, by using simulations, that if the elimination rates are evaluated using the two sample method, the error is actually lower than when using the multiple sample and curve fitting approach – a rare statistical example of fewer samples providing greater precision. By estimating the final pool size from body water and using the two sample method, this source of error can be ignored. The second assumption is that the materials leaving the body take isotopes with them at the same enrichment as body water. This assumption is violated because of physical isotopic fractionation events that
occur when molecules change phase. Evaporating water, for example, carries with it slightly less deuterium and oxygen-18 than the water left behind. Because we know the extent of fractionation from in vitro studies, the equations used to evaluate CO$_2$ production can be modified to account for this effect. However, all the parameters necessary to make these corrections are not fully known and so several different equations are available – the principal ones for small animals being the Lifson & McClintock (1966) equation 35, the Nagy (1980) equation 2, and the Speakman (1997) equation 7.17. Validations of the method post-1997 have confirmed the last of these as being the most accurate (Visser & Schekkerman 1999; Visser, Boon & Meijer 2000), primarily because this latter equation makes a more realistic assumption of evaporation (25% of total water loss) compared with 50% in Lifson & McClintock (1966) and 0% in Nagy (1980).

The third assumption is that the isotopes of hydrogen and oxygen only take part in reactions that involve water and CO$_2$. There are two types of error that could be generated here. First, the isotopes may combine with other substrates and leave the body – this would increase the apparent elimination rate. Alternatively, the isotopes may exchange with substances in the body, thus increasing the size of the apparent pool in which the isotopes are turning over. The most significant problems occur with hydrogen, which is involved in both types of reaction. Labelled hydrogen exchanges reversibly with hydrogen on the exposed amino groups of proteins (Culebras & Moore 1977; Matthews & Gilker 1995) and is involved in irreversible incorporation into lipids during de novo lipogenesis – indeed this latter effect is an accepted method for quantification of lipid synthesis (Guo et al. 2000). Because of the reversible exchange, estimates of the body water pool size based on hydrogen isotope dilution are around 3–4% greater than the oxygen space (Coward et al. 1985; Schoeller et al. 1986; Prentice 1990; Speakman, Nair & Goran 1993; Coward, Ritz & Cole 1994; Racette et al. 1994). There are consequently two ways to address this problem – the first is to modify the equation so that each turnover is expressed relative to its own dilution space (so called two pool models) and the other method is simply to ignore the discrepancy and use the oxygen dilution space as a true estimate of the body water pool (single pool models). Deciding between these approaches depends on the amount of hydrogen that is also involved in the irreversible reactions (Speakman 1987), which in practice comes down to an issue of body size. Our best estimates are that for animals weighing less than 4 kg, the single pool approach works best, and for larger animals the two pool models work best (Speakman et al. 2001a).

Essentially the consequences of this assumption can therefore be eliminated (or at least minimized) by appropriate choice of equation.

For a long time this difference in equations posed no problem. Studies of DLW until the late 1980s were largely performed by researchers looking at small animals using appropriately the single pool model and researchers studying Humans appropriately using the two pool model. As interests of the different laboratories have expanded however, they have tended to take with them the methodologies established in their own areas and hence we now have laboratories formerly working on small animals applying the technique to large animals but still using the small animal equation, and conversely laboratories formerly studying only Humans applying the technique to small animals but using the large animal (Human) equation (reviewed in Speakman 1993, 1997). The calculations using the different equations produce answers that differ by between 2 and 30%, depending on the actual isotope turnovers involved, so this is not a trivial problem. Fortunately there are signs that the need to use the appropriate equation is becoming recognized by the wider community (e.g. Costa & Gales 2003).

Probably the most significant assumptions therefore are the final two, since these are not accounted for in current equations. The fourth assumption is that once eliminated, isotopes do not re-enter the body, and unlabelled CO$_2$ does not enter the body, creating a CO$_2$ turnover that would be detected as CO$_2$ loss. The conditions that would generate sufficient isotope re-entry to be an issue are generally unrealistic and can probably be ignored. However, there are probably many circumstances where subjects may inhale elevated levels of CO$_2$ derived from other animals or the environment – for example mothers suckling their young and animals that live communally in restricted spaces such as burrows. Because inspired CO$_2$ exchanges with body water before re-expiration, it contributes to the estimates of elimination. Nagy (1980) forced Kangaroo Rats (Dipodomys sp.) to inhale artificially high CO$_2$ continuously during a validation study and found the overestimate in CO$_2$ production that was about 80%. Poppitt, Speakman & Racey (1993) evaluated the energy demands of Suckling Shrews (Sorex arenes) and concluded that isotope exchange from the pups might have compromised the estimated energy expenditure estimate compared with food intake measures. Apart from these studies, however, our knowledge of this potential error is rudimentary.

Finally, an assumption not documented by Lifson & McClintock (1966), Mullen (1973) and Nagy (1980) is that the background isotope enrichment is constant during the measurement period. The reason they ignored this was probably not ignorance, but because this is another assumption the importance of which depends critically on body size – and in small animals it is generally irrelevant. This is because small animals are generally enriched to much higher levels than large animals because of cost considerations (see below), and they are recaptured and sampled over much shorter durations – both of which minimize the impact of background variability. Measurements of day-to-day background isotopic variation in Humans, however (Horvitz & Schoeller 2001), suggest that the assumption
of constant background enrichment ultimately limits the accuracy of the technique, so that further improvements in mass spectrometric technology will not enable us to reduce the doses or extend the durations of experiments.

HISTORY
The discovery of isotopes revolutionized the study of physiology, because rare isotopes, having almost identical physico-chemical properties to the native isotopes, are almost perfect tracers. Lifson's pioneering experiments using oxygen-18 were performed in the late 1940s and in 1955 Lifson et al. published the seminal study using both oxygen-18 and deuterium as a potential method to trace CO₂ production. For over 25 years the technique was utilized only sporadically, principally on small birds, reptiles and mammals. Lifson et al. (1975) suggested that a DLW study of a Human at the same doses used in small animals would cost around US$4000 (equivalent at 2003 prices to around US$150 000). Yet advances in mass spectrometry technology meant the required dose progressively declined and the first Human measurements were made in the early 1980s (Schoeller & van Santen 1982). Throughout the 1980s and 1990s, utilization of the method increased exponentially (Prentice 1990; Speakman 1997). This expansion in use was particularly strong in the study of Human energy metabolism where it has become regarded as a 'gold standard' technique. Applications in ecology have not expanded at quite the same rate. In the early part of the new millennium, the numbers of published DLW studies have stabilized at around 120 per annum, which is slightly down on the peak of almost 200 per annum in the late 1990s.

VALIDATIONS (PRECISION AND ACCURACY)
Accuracy is the closeness of a measurement or computed value to its true value, whereas precision is the closeness of repeated measurements of the same quantity to each other. Thus, a precise measurement is not necessarily an accurate one. Validation experiments give an indication of the overall accuracy of the method by comparing estimates of MR with simultaneously obtained measurements using, for example, indirect calorimetry. Validations have been performed on a wide range of animals weighing between 300 mg (Bumble Bees *Bombus terrestris*; Wolf et al. 1996) and Humans (80 kg) (reviewed in Speakman 1997 who included all validations up to 1997). Several validations since that time have extended the range of conditions studied to include growing animals (Visser & Schekkerman 1999; Visser et al. 2000) but have not extended the size range. There is a generalized feeling in the field that the technique has been validated on so many occasions now that further validation studies are not a prerequisite for applications of the method. This confidence may be misplaced given the possibility that assumptions not accounted for in the calculation may still be violated in certain conditions – particularly isotope re-entry or rebreathing unlabelled CO₂. Some caution, therefore, needs to be applied. Moreover, validations would enable the most appropriate model/equation to be selected. Validation studies have generally included sample sizes of around 6–12 individual animals. In almost all the validations performed to date using these sample sizes, the average discrepancy between the DLW and the reference method of indirect calorimetry has been less than 10%. Across all validations, the mean discrepancy in the sample of studies reviewed by Speakman (1997) was 3·1% in mammals, 2·4% in birds and 0·5% in reptiles.

There have been few validation studies in invertebrates. Early studies (King & Hadley 1979; Cooper 1983) on scorpions and tenebroid beetles indicated large errors when the method was applied to arthropods. This might be anticipated because the penetration of tracheoles direct to cells facilitates gas transfer but would minimize the time that respiratory CO₂ would have to come to isotopic equilibrium with body water, which is obviously necessary for the method to work. More recent validation studies in insects, however (Wolf et al. 1996), have suggested errors in the same range as those observed in vertebrates, suggesting the early problem may only have reflected difficulties handing small samples in the mass spectrometers available 20–25 years ago. Clearly, the method generally provides an accurate evaluation of the mean energy demands of a group of individual vertebrates and perhaps also invertebrates as well – although this latter point needs more verification.

For individual estimates however, the estimates from DLW can be substantially more deviant from the reference method. It is not unusual within a group of 10 individuals under validation to find some individuals with estimated CO₂ productions that differ by more than 20% from the reference measure. The reasons for these discrepancies remain uncertain. Few studies have estimated precision of DLW measurements (e.g. Ricklefs, Roby & Williams 1986). Speakman (1995) presented techniques that would allow derivation of the analytical precision of individual DLW estimates. This technique takes advantage of the fact that the method depends on replicated isotope determinations made at several time points. By exhaustively iterating the individual values in this calculation, a distribution of estimates can be derived and a precision error assigned to each individual estimate. The obvious question is whether discrepancies between DLW and observed CO₂ productions can be linked to individual variability in precision of the DLW estimates. By reanalysing data from a validation study on bats (Speakman & Racey 1988) and using the derived individual precisions for each animal, Speakman (1995) concluded that individual discrepancies from the reference method were not completely explained by variations in analytical precision. This raises an interesting and as yet unanswered question. If an individual animal deviates from the reference method by more than the analytical precision, is this trait connected to that particular animal or is
this a random deviation with time? The answer to this question is important, because if it is a trait connected to the individual, this would indicate that some individuals have physiological characteristics that cause the deviations. If we can understand what these characteristics are, we could potentially measure and correct for them on an individual basis. This would enable us to refine the method so that individual estimates using the technique would become usable.

Data from these previous studies suggest that the differences observed between DLW estimates and indirect calorimetry cannot be explained solely by variability in precision. This variability in precision can be modelled using the techniques presented by Speakman (1995) but, as this author acknowledged, this still begs the question as to what causes the additional discrepancy between predictions and measurements. Only by attempting to understand and model this further variation can the accuracy of our DLW estimates be assessed and enable the mean estimate to be presented with an appropriate standard error. Most current experimental protocols do not present sufficient data and information to enable such a model to be constructed. Calibration or validation of DLW against indirect calorimetry in a laboratory context for each species to be studied would allow an assessment of the accuracy of the technique to be made. However, such a procedure would be logarithically intensive, requiring \( V_O \) and \( V_CO_2 \) to be measured and estimated from several animals, over the course of several days (according to the size of the study species) and preferably at a range of activity levels to match those found in the field (see, for example, Speakman & Racey 1988). The standard error of the mean estimate can then be determined as described in Zar (1984). It is probably for these logistical reasons that most studies have tended to ignore this potentially important component of variability. However, as a result of this, conclusions drawn from comparisons of DLW estimates with a standard error of the mean calculated from each individual estimate on the assumption that it has a high degree of accuracy, should be regarded with caution.

**ADVANTAGES**

The primary advantage of the DLW method is that it provides a direct estimate of CO\(_2\) production (and hence FMR) that is independent of assumptions concerning the mode in which the utilization of energy has been made. The composite nature of the final estimate is a significant advantage when compared with time and energy budgeting approaches to estimation of FMR, because the myriad of combinations of factors that might impact on energy expenditure is such that assigning accurate estimates of energy expenditure to different components of the time budget becomes virtually impossible.

A second advantage is the ease with which the method can be applied in the field. In its simplest form, the technique requires only that an animal be captured twice, and during these captures injected once and bled twice. Indeed, there are protocols available for the method that utilize knowledge of the isotopic enrichment of the injection solution to dispense with even the first blood sample – the single sample approach (Ricklefs et al. 1986; Webster & Weathers 1989; Williams 1993; Williams & Dwinnell 1990).

Several studies have been performed to evaluate the impact of the method on animal behaviour (reviewed in Speakman 1997) and the general consensus of these studies is that the two-sample method does not have significant behavioural impacts that might compromise the estimates of daily energy metabolism. More recent studies have suggested that there are some minor behavioural impacts of the two-sample technique – but even these are eliminated when using the single sample approach. Yet even these impacts may be eliminated in the near future as researchers find less invasive ways to dose and sample their subjects. For example, Anava et al. (2002) dosed Arabian Babblers (Turdoides squamiceps) by injecting dead items of their insect prey with the isotope and then presenting this food in an area where the bird would ingest it. Using a single sample approach, they then needed only to capture the animal at the end of the measurement period for a final blood sample. Even this final sample may become redundant in some species, since Haggarty et al. (1998) measured the energy demands of Red Deer (Cervus elephas) by extracting water from droppings that they had observed the focal animal producing. However, this completely non-invasive application of the method will not always be feasible – for example in a recent study of Meerkats (Suricata suricatta, Scantlebury et al. 2002), the authors attempted to measure final isotope enrichments using urines deposited in dry sand collected immediately after production, but found them too variable to provide reliable data. Nevertheless, as the volumes of sample required to make isotope enrichment measures continues to fall, new possibilities open up. One we are currently actively pursuing is the possibility of dispensing with the blood samples and replacing them instead by samples of breath in which we can measure oxygen-18 enrichment in exhaled CO\(_2\) and deuterium levels in trapped water vapour. This has already been achieved for CO\(_2\) measures of oxygen-18 (Krol & Speakman 1999).

Finally, a major advantage of the method is the size of the animals on which it can be used. Currently the smallest animals on which estimates of energy metabolism have been published are Bumble Bees weighing around 300 mg (Wolf et al. 1996). Judicious utilization of the single sample approach (Ricklefs et al. 1986; Webster & Weathers 1989) and modern mass spectrometric methods for sample preparation mean that this lower limit may be extended downwards in the next 5–10 years by one, two or even three orders of magnitude. A 3-mg insect probably contains 2 \( \mu \)l extractable water which is already analysable by commercially available pyrolysis machines, and a 0.3-mg animal might provide
200 nl of water. Generation of isotope enrichments from 50 nl volumes is already technically within sight (Morrison et al. 2001).

**DISADVANTAGES**

A major disadvantage of the method is that the accuracy of individual measures is such that it cannot be reliably used to estimate the energy demands of a focal individual subject. This in turn will affect the standard error of the estimate of the mean value from a group of animals. As yet, the causes of the individual discrepancies remain obscure, but this is an active area of work and developments in the next decade may make individual-based estimates feasible. Therefore, the technique can be used at present only to provide an estimated MR of a group of individuals (Speakman 1998). Calculations based on the error in validation studies indicate that the mean derived from a group of nine or ten individuals would have a mean accuracy error of about 2–3% on the estimated CO₂ production.

CO₂ production, however, is not energy expenditure and a second disadvantage of the method is that while it provides a direct estimate of CO₂ production, a significant error could be generated if this was converted to energy using an inappropriate RQ. At present our only solution to this problem is to utilize a best guess RQ based on dietary information (i.e. a food quotient, FQ). Food quotients are derived from knowledge of the macronutrient compositions of the diet. Since in steady state, non-growing animals oxidize all the food they ingest, the average respiratory quotient should reflect the macronutrients in the food. The food quotient is the projected RQ for complete oxidation of a given food type. Although studies in Humans have indicated use of FQ is as good as RQ for converting CO₂ production to energy demands (Black, Prentice & Coward 1986) there are some obvious limitations when considering its use in wild animals – primarily they may be growing and we have major uncertainties about their precise dietary composition. Two innovative methods have been proposed to estimate RQ using isotopes (Speakman & Racey 1987; Haggarty, McGaw & Franklin 1988), but as yet, neither method has been well validated.

The third disadvantage of the DLW method is its cost. In the past 5 years, the cost of the oxygen-18 isotope has more than doubled because of the rapidly expanding demand for Positron Emission Tomography (PET) scans in the USA. These scans use oxygen-18 as a reagent, but the limited number of suppliers (Isotec and Cambridge Isotopes in the USA and Rotem in Israel) has led to competition and driven prices up. Furthermore, the advances in techniques to reduce the dose of oxygen-18 required in DLW experiments are unlikely to progress much further and therefore the current cost of isotopes of approximately $1000 for a 70 kg Human is unlikely to decrease. For smaller animals, the costs of isotopes become trivial, but the costs of analysis become significant. Typical costs in 2003 for a two sample estimate of FMR are around US$150–200 per animal. Single sample estimates are not quite half this because of the costs of running quality control standards.

When studying free-living animals, there is also the potential problem of recapturing animals within given time windows. Even if an animal is ultimately recaptured, if all the isotopes have been eliminated, an estimate of CO₂ production is not possible. This lost isotope adds to the costs of any study. In our experience, many DLW studies achieve marginal or inadequate sample sizes because of the failure to plan adequately for failed recaptures (and losses at the processing stage of analysis). Depending on the questions being asked, another potential disadvantage of the method is that, because the estimate is an integrated composite estimate of energy demands, it yields no breakdown of the component costs that contribute to the total. Although certainly not all studies require this information, a good many ecological enquiries do. To overcome this problem, Flint & Nagy (1984) used estimates of FMR combined with independently observed time budgets of Sooty Terns (Sterna fuscata) to estimate the energy demands of flight using regression of the FMR against the percentage of time spent in flight, and then extrapolating this to 100% to estimate the flight costs. Utilizing the flight cost estimate with the time budget then allows some breakdown of the component costs. This method has been used subsequently to evaluate flight energetic costs of other animals, e.g. Long Eared Bats (Plecotus auritus) (Racey & Speakman 1987) as well as costs of swimming/diving in seals (e.g. Arnould, Boyd & Speakman 1996; Costa & Gales 2003). To work effectively, this method requires much additional data collection in addition to the DLW measurements. In theory, however, one might imagine that in a large enough data set, this general approach could be extended in a multiple regression analysis to elucidate costs of several components and then reconstruction of more complex time–energy budgets. However, this approach depends critically on the assumption that behaviours occur independently of each other – which will seldom be the case. The problems of covariance in behaviours for this method have been discussed in detail in Speakman (1997).

Finally, because the method relies on distinguishing the elimination curves of oxygen-18 and deuterium, the error in the method increases dramatically as the ratio of CO₂ production to water production gets smaller (high water production relative to CO₂ production). In practice this makes applications to ectothermic aquatic organisms impossible, and there may also be problems with aquatic air-breathers when in water. For example, average FMR of Antarctic Fur Seals, Arctocephalus gazella, foraging at sea over several days and estimated by DLW (Costa, Croxall & Duck 1989) was similar to the maximum MR that it was possible to elicit from California Sea Lions, Zalophus californianus, swimming in a water channel (Butler et al. 1992). Thus, either free-ranging Fur Seals are capable of raising their MR to a substantially higher level than
Sea Lions in the laboratory, or the DLW method overestimates their FMR when at sea.

Heart rate method

**PRINCIPLES AND ASSUMPTIONS**

The heart rate \( f_H \) method for estimating \( \dot{V}_O_2 \), and hence FMR, is based on Fick’s convection equation for the cardiovascular system:

\[
\dot{V}_O_2 = f_H \times V_s (C_aO_2 - C_vO_2)
\]

where \( V_s \) is cardiac stroke volume – the amount of blood pumped per heart beat; \( C_aO_2 \) is oxygen content of arterial blood and \( C_vO_2 \) is oxygen content of mixed venous blood. This method relies on the premise that a change in \( f_H \) is a major component in the response of the cardiovascular system of a species to an increase in the demand for oxygen. However, Stevens & Randall (1967) concluded for Rainbow Trout, *Oncorhynchus mykiss*, during moderate exercise, that an increase in oxygen delivery to the tissues is largely the result of changes in \( V_s \) with only small changes in \( f_H \) and a similar situation was also found for Atlantic Cod, *Gadus morhua* (Webber, Boutilier & Kerr 1998). More recent studies on Rainbow Trout have not supported the conclusion of Stevens & Randall (1967) (Altimiras & Larsen 2000; Brodeur, Dixon & McKinley 2001) and \( f_H \) changes with \( \dot{V}_O_2 \) in Pike, *Esox lucius* (Armstrong 1986), Atlantic Salmon, *Salmo salar* (Lucas 1994) and Brown Trout *Salmo trutta* (Beaumont, Butler & Taylor 2003) as well as in members of the other Classes of vertebrates, except perhaps in association with the increase in \( \dot{V}_O_2 \) after feeding (McPhee et al. 2003).

Thus, if the term \( V_s (C_aO_2 - C_vO_2) \), which is known as the oxygen pulse (OP) – the amount of oxygen consumed by the animal per heartbeat (Henderson & Prince 1914) – is constant or changes in a systematic fashion, there will be a linear relationship between \( \dot{V}_O_2 \) and \( f_H \) (Fig. 2), so that the latter could be used to determine the former (see Butler 1993). It is clear from Fig. 2 that if the relationship passes through zero, then OP is constant. This is unlikely to occur in a real animal, as at least \( (C_aO_2 - C_vO_2) \) will most probably increase during increased activity (see, for example, Butler, West & Jones 1977; Butler et al. 1993), which means that OP is not normally constant and the relationship does not pass through zero. If only one calibration curve of \( \dot{V}_O_2 \) vs \( f_H \) is produced, then the assumption is that the individual components of the oxygen pulse will maintain a reasonably constant relationship with \( \dot{V}_O_2 \) throughout an animal’s annual cycle, but this may not be the case. Any change in the size of the heart, which could affect \( V_s \), and any factor which could affect \( (C_aO_2 - C_vO_2) \) could alter the relationship between \( f_H \) and \( \dot{V}_O_2 \). For example, if the heart, and hence \( V_s \), became larger as a result of an improvement in physical fitness, then \( f_H \) could be lower for a given \( \dot{V}_O_2 \). Also, a thermally stressed bird or mammal is likely to increase peripheral blood flow in order to increase heat loss. This may well be achieved, at least in part, by an increase in \( f_H \), but without an accompanying increase in \( \dot{V}_O_2 \). A similar change in the relationship between \( f_H \) and \( \dot{V}_O_2 \) is likely to occur in animals exposed to a hypoxic environment (most usually high altitude in air-breathing animals).

**HISTORY AND DEVELOPMENT OF THE METHOD**

Boothby (1915) and Krogh & Lindhard (1917) demonstrated that there is a linear relationship between \( f_H \) and \( \dot{V}_O_2 \) in Humans during exercise, while Henderson, Haggard & Dolley (1927) and Bock et al. (1928) showed that the relationship is affected by the level of physical fitness of the subjects. According to Lundgren (1946), Berggren suggested in 1945 that \( f_H \) could be used to determine ‘energy output during athletics’. To quote from Lundgren (1946), ‘For this purpose the subject is calibrated by constructing a pulse-oxygen consumption diagram from treadmill or bicycle ergometer data. From this diagram it is then possible to read the oxygen consumption corresponding to the pulse-rate recorded during athletics’. Lundgren (1946) himself used \( f_H \) to estimate \( \dot{V}_O_2 \) during wood-cutting by lumber workers. He also determined the accuracy of his estimates by comparing them with values for \( \dot{V}_O_2 \) that were measured by respirometry. The overall difference was –1.17% and Lundgren concluded that pulse-rate is suitable for calculating metabolic rate during industrial work. A similar conclusion was reached by Malhotra, Sen Gupta & Rai (1963), who reported average errors of +1.7% to +3.7%, depending on whether \( f_H \) was below or above 95 beats min\(^{-1}\), respectively. However, since then, there have been many studies on Humans and it has often been concluded that there are large individual errors.

![Fig. 2. Hypothetical data to illustrate the changes in oxygen pulse (*) that are required to produce a linear relationship between rate of oxygen consumption and heart rate when that relationship passes through zero (-) and when it does not (--), which is invariably the case (Butler 1993).](image-url)
associated with the $f_H$ method and that it is not recommended for estimating mean daily energy expenditure (e.g. Washburn & Montoye 1986). Improvements on the predictive power were obtained if heart rates were categorized as resting or active (Livingstone et al. 1990) with a single resting metabolic rate being assigned to the former and the latter producing a $f_H/V_o$ relationship. Psychological stress and mental tests can cause an increased or ‘additional’ heart rate above that predicted by the $f_H/V_o$ relationship obtained from graded exercise, in both resting and active Humans (Blix, Strømme & Ursin 1974; Carroll, Turner & Prasad 1986) and different $f_H/V_o$ relationships are obtained for arm exercise and leg exercise (Vokac et al. 1975). Furthermore, an $f_H/V_o$ relationship obtained during dynamic exercise, such as running, is different from that obtained during static exercise, such as holding a heavy object (Maas et al. 1989).

Statistically significant relationships between $f_H$ and $V_o$ have been demonstrated in several species of birds and mammals (Owen 1969; Bamford & Maloiy 1980; Pauls 1980; Nilssen et al. 1984; Butler et al. 1992; Nolet et al. 1992; Williams, Friedl & Haun 1993; Bevan et al. 1994, 1995b; Hawkins et al. 2000; McCarron et al. 2001; Ward et al. 2002). This includes diving birds and mammals, and Fedak (1986) made the important observation that, if $f_H$ and $V_o$ are averaged over a complete dive cycle, there is a significant relationship between $f_H$ and $V_o$. Heart rate has also been successfully used as a qualitative indicator of field metabolic rate of birds while flying under different conditions (Weimerskirch et al. 2000, 2001).

CALIBRATION METHODS

Most of the relationships in the above studies on birds and mammals were obtained during some form of increasing work load, such as running or swimming, but some studies (e.g. Webster 1967; Owen 1969; Wooley & Owen 1977; Flynn & Gessaman 1979; Gessaman 1980) used exposure to low ambient temperature to raise $V_o$ and $f_H$ in resting animals. Whether or not these relationships are similar to those that would be obtained during exercise in these species remains to be seen, but Froget et al. (2002) found that, for the King Penguin, Aptenodytes patagonicus, the relationship obtained during exposure to the cold is significantly different from that obtained during exercise (Fig. 3). The pectoral muscles of volant birds are larger than the leg muscles (Butler 1991) and Ward et al. (2002), reported a clear difference in the relationships between $f_H$ and $V_o$, obtained from running and flying geese, with the latter being much steeper (Fig. 4). However, in this case, the range of $f_H$ while running and flying did not overlap, suggesting that a single curvilinear relationship may link $f_H$ and $V_o$ across all types of exercise. In Gentoo Penguins, Pygoscelis papua, running or swimming underwater (i.e. using their flippers, which are modified wings), the $f_H/V_o$ relationships were not significantly different from each other (Bevan et al. 1995b), but in this case the range of $f_H$ while running and swimming overlapped considerably. Despite the latter finding, it is strongly recommended that calibrations are carried out under conditions which simulate, as closely as possible, those that the animal will experience in its natural environment, especially with respect to the way in which metabolic rate will be raised above resting. It is also important that the range of $f_H$ during the calibration procedure covers that obtained from animals in the field. The heart rate of captive Barnacle Geese, Branta leucopsis, flying in a wind tunnel was significantly higher than that of migrating wild geese (Butler, Woakes & Bishop 1998; Ward et al. 2002), and this suggests that the animals used for calibration should be in a similar physiological state to those to be monitored in the wild.

Having established that there is a significant relationship between the two variables, it is evident from the data obtained from Humans that it is necessary to determine whether or not the relationship is different under different conditions or at different times of the year, etc. For captive Barnacle Geese at least, there was no difference in calibrations carried out on different individuals, 10 years apart (Nolet et al. 1992; Ward et al. 2002), while for Tufted Ducks, Aythya fuligula, $f_H$ was a consistent predictor of $V_o$ in two sets of experiments on different animals, separated by 9 years (Woakes & Butler 1983; Bevan & Butler 1992). In addition, for Macaroni Penguins, Eudyptes chrysolophus, there was no difference in the calibrations of two groups of females which were either active/breeding or inactive/moulting (Green et al. 2001). However, the situation is more complex during the breeding period of the King Penguin. This species alternates periods ashore incubating or looking after the chick with many days at sea, while its partner does the opposite. Thus, during the periods ashore, the birds are fasting and...
In lower vertebrates, of course, body temperature will vary on a diurnal and/or on a seasonal basis, so in these animals the effect of body temperature on the $f_H/V_O_2$ relationship must be taken into account. This has been achieved with three species of fish, Pike (Armstrong 1986), Atlantic Salmon (Lucas 1994) and Atlantic Cod (Webber et al. 1998) and a reptile, the Galapagos Marine Iguana Amblyrhynus cristatus (Butler et al. 2002), at a similar range of temperatures that these animals experience in the field. In Pike and Atlantic Salmon, a single linear relationship between $f_H$ and $V_O_2$ exists up to a $f_H$ of 55–60 beats min$^{-1}$, although there may be a difference between the sexes in the salmon, whereas in the cod and iguana the effect of temperature is to vary the intercept of the relationship between the two variables, rather than to extend a single regression line (Fig. 5).

**VALIDATIONS**

In addition to constructing calibration curves, it is also desirable to perform validation experiments to assess the probable accuracy of any estimates of $V_O_2$ obtained in the field and, in the case of situations such as the breeding King Penguins, to determine which method of estimating $V_O_2$ is most accurate. Validations have been performed by using indirect calorimetry (respirometry) as the ‘gold standard’. The accuracy of the measurement of $V_O_2$ and $V_CO_2$ can be assessed by the method described by Fedak, Rome & Seeherman (1981), but converting $V_O_2$ to MR is not always straightforward (see Simonson & DeFronzo 1990).

An animal is placed inside a respirometer on a treadmill and $V_O_2$ (and $V_CO_2$, in order to determine RQ) and $f_H$ are continuously measured over a period of 2–3 days. The animal is randomly exercised at different levels and rested during the day and at night. It is then possible to compare $V_{O_2}$ estimated from $f_H$ ($V_{O_2}$ est) with that measured directly over the whole validation period ($V_{O_2}$ meas) and for different times during the day or night and for different levels of activity. Several important points have emerged from the validation experiments that have been performed on different species of animals.

There is a relatively wide range in the discrepancy between $V_{O_2}$ meas and $V_{O_2}$ est (error) between individual animals (overall range, −25·9% to 26·5%; Nolet et al. 1992; Bevan et al. 1994, 1995b; Hawkins et al. 2000; Froget et al. 2001), which means that the $f_H$ method cannot be used to estimate the FMR of individual animals unless, perhaps, each animal is individually calibrated. This is normally logistically not possible. However, the above studies also demonstrated that the mean $V_{O_2}$ est for a number of animals was, on average, within a few per cent (range, 3·7% to −2·13%) of the mean $V_{O_2}$ means for the same individuals. Lucas & Armstrong (1991) have demonstrated that heart rate provides an accurate estimate of apparent specific dynamic action and meal size for Pike, independent of the mass of the fish, meal size and environmental temperature.
One benefit of using the heart rate as an indicator of FMR is that the calibration data giving the relationship between \( f_{IH} \) and \( V_O \) are also used in the calculation of the SE of the estimate (SEE) of \( V_O \) in the field using mean \( f_{IH} \) from the animals in the field. How SEE is calculated will depend on the nature of the variation between calibration animals, as described by Zar (1984), Hawkins et al. (2000) and Green et al. (2001). In addition, such analysis can also be used to identify the minimum number of animals from which data would need to be obtained during calibration and from the field in order to obtain an acceptably small SEE. In both cases, approximately eight animals seemed to be appropriate for Macaroni Penguins.

**Use in the field**

In earlier studies on free-ranging animals, \( f_{IH} \) was obtained by telemetry, which is limited by the range of the transmitter and proximity of the receiver (e.g. Thompson & Fedak 1993). However, the widespread use of the \( f_{IH} \) method in the field has been made possible by the development of data loggers that can store average heart rate for many weeks (Woakes, Butler & Bevan 1995), and now for over a year. The principal advantage of the method is that it can provide estimates of MR at a far greater temporal resolution than other methods; the resolution being limited only by the calibration procedure and the method of recording \( f_{IH} \). Also, being implanted, the data logger has no effect on the aero- or hydro-dynamics of flying or aquatic animals. In conjunction with behavioural data, the energy costs of individual activities can be determined. Unfortunately, though, many of the behavioural devices, such as satellite transmitters (Butler et al. 1998), salt-water indicators (Bevan et al. 1995a), speedometers (Boyd et al. 1999) and time–depth recorders (Bevan et al. 1997) are attached externally and hence lead to increased drag of flying and aquatic animals (Bannasch, Wilson & Culik 1994). For time–depth data, however, the ‘Woakes’ data loggers now incorporate a pressure sensor (Green et al. 2002, 2003). Although there is a period of about 2 days after implantation of a data logger when the animals show slight effects of the procedure (Bevan et al. 1995b), there is no evidence of any effect on survival or on overall behaviour of Black-Browed Albatrosses, Diomedea melanophris (Bevan et al. 1995a), Gentoo Penguins (Bevan et al. 2002), Macaroni Penguins (Green et al. 2004b), Common Eider Ducks, Somateria mollissima (Guillemette et al. 2002) or Barnacle Geese (P. Butler et al. unpublished data).

Field data using the \( f_{IH} \) method have been obtained from Pike (Lucas et al. 1991), Antarctic Fur Seals (Boyd et al. 1999), Gentoo Penguins (Bevan et al. 2002), Macaroni Penguins (Green et al. 2002, 2003, 2004a), and King Penguins (Froget et al. 2004). An interesting feature of the studies on aquatic birds and mammals is that the estimates for MR when the animals are ashore are similar to those obtained by other workers using the DLW method, but those obtained from the animals when they are at sea are between 40% and 60% lower than those using DLW. This is consistent with the outcome of a validation study of the two methods carried out on California Sea Lions in a water channel (Boyd et al. 1995). This means that when animals are at sea, the \( f_{IH} \) method may underestimate FMR and/or the DLW method may give an overestimate (see DLW section).

**Advantages**

A major advantage of the heart rate method is that it is able to provide estimates of the metabolic costs of specific activities over long periods lasting up to a year using current implantable data loggers. In addition, the loggers are usually able to monitor one or more other physiological or behavioural variables such as body temperature, dive depth, acceleration and body position (attitude), thus allowing details of the metabolic rates to be correlated to other features of the animal’s behaviour and environment. The current capacity of modern \( f_{IH} \) data loggers allows us to obtain minute by minute ‘pictures’ of what animals are doing for periods extending for many months, thus providing an unparalleled insight into the world of individual animals.

Another advantage is that there is no limited time period within which the animal has to be recaptured. Once stored, data in a non-volatile memory will be available for downloading at any subsequent time. Therefore, once the recording system has been applied to the animal, it can be recovered at a time that is convenient for capturing the animal, for example when an aquatic bird or mammal comes ashore to breed or when a wild bird is in moult and cannot fly, and/or at the convenience of the investigator. Animals and implanted loggers survive well; a Barnacle Goose was recaptured 5 years after implantation and there were good data on the logger (A.J. Woakes & P.J. Butler, unpublished observation).

Although it is not possible to use the heart rate method to determine FMR from individual animals, it is possible to determine the SE of the estimate of the mean value obtained from a number of individuals, so that valid statistical tests can be applied to mean values obtained from the same animals under different conditions or from a different group of animals. The method is probably not restricted to vertebrates (see, for example Hamilton & Houlihan 1992; working on the European Shore Crab Carcinus maenas). Moreover, its use within the vertebrates is not restricted in the same way as DLW, which cannot be applied to animals where there is excessive water turnover like fish and, possibly, air-breathing aquatic vertebrates when in water. Finally, because measures of heart rate result in estimates of oxygen consumption, there is a smaller error when these are converted to estimates of energy expenditure and RQ is unknown than for the DLW method, which derives an estimate of \( CO_2 \) production.
The most significant problem with the $f_H$ method is the fact that for each new species that is studied it is necessary to derive a calibration equation. Indeed, even for species where a calibration equation has been previously constructed, further calibration may be necessary if the study includes novel situations that might affect the relation between $f_H$ and $V_O$. The logistical difficulties of performing such calibration work are probably the most significant constraints on current applications. At present there are about 10 papers per year published using this method, compared with 12 times that number using DLW. Above a certain animal body size the logistical difficulty of performing a calibration becomes the most important limiting factor on applications of the method. At present, the largest animals on which the technique has been used have a mass of approximately 350 kg (e.g. Brosh et al. 1998).

Another possible disadvantage is that, to ensure there is little or no effect on the wild animal to be studied, the recording system should be surgically implanted and subsequently removed.

The size of the most recent ‘Woakes’ implantable $f_H$ data loggers ($40 \times 30 \times 13$ mm$^3$ and 20 g) places a lower limit on the size of animal that can currently be studied at approximately 1 kg. Smaller, non-logging transmitters are available that can be monitored externally, and are cheaper than the logging devices, costing around US$250, but these require continuous access to the animals to record data. Such transmitters weigh around 1 g, placing the lower size limit over an order of magnitude smaller at less than 40 g (Cochran & Wikelski 2004). Below this size, the $f_H$ method cannot be used at present. The major limitation on mass of a transmitter or data logger is the battery, and apart from some radical developments in battery or power supply technology, it is unlikely that this lower limit will be significantly extended downwards in the near future.

The heart rate method is a relatively expensive technique. The current cost of a particular implantable data logger is approximately US$750, with data from $c. 6–10$ animals necessary for a successful study. Excluding the costs associated with any validation/calibration work which might themselves be significant, and the one-off costs of equipment to download the data from the loggers, and assuming a reuse rate of loggers of 50%, a typical study would cost around US$3000–3750. Unlike the DLW method, these costs are independent of the size of the animal under study (see Fig. 6 for comparative costs of both methods).

The success of a study using the heart rate method depends on the reliability of the recording device. If attached externally, it may become detached and lost. If implanted (or attached externally to aquatic animals), the waterproofing may be unable to prevent moisture from reaching the electronics, which will cause a short circuit and reduce the life of the battery. This is more likely to be a problem with deeply diving aquatic animals where the hydrostatic pressure could regularly vary over several tens of atmospheres. However, year-round data have been successfully obtained from Macaroni Penguins for which 90% of the dives were down to 35 m (Green et al. 2003a), suggesting this is not as serious a problem as it might first appear.

**Conclusions**

Even though neither the DLW nor the $f_H$ method is perfect, both can be used to determine the rates of CO$_2$ production or O$_2$ consumption, and hence MR, of animals in the field. Despite their apparent limitations, it is important to recognize that at present these methods are all we have. Validation studies for both techniques have indicated that estimates of the mean MR of a number of individuals are usually within less than 5% of that measured directly, i.e. on average they are acceptably accurate. Validation studies also indicate that at present neither technique is sufficiently accurate that we can place confidence in individually derived estimates. The utility of each method will depend, to a large extent, on the question being asked. A summary of the strengths and weaknesses of the two methods is presented in Table 1. Overall the balance of evidence suggests that $f_H$ will be a superior approach for larger (>1 kg) animals, and where a detailed breakdown of costs is needed and DLW will be superior for smaller (<1 kg) animals and where a detailed breakdown is not required. For the smallest animals, only DLW is usable.
Table 1. A summary assessment of the relative strengths and weaknesses of the DLW and heart rate methods of measuring field metabolic rate. Scores are subjective but are meant as a guide to strengths and weaknesses. The ratings are on a system of five stars representing a particular strength of the method and one star representing a particular weakness.

<table>
<thead>
<tr>
<th></th>
<th>Doubly labelled water</th>
<th>Heart rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Useful in animals &gt; 100 kg</td>
<td>**</td>
<td>*****</td>
</tr>
<tr>
<td>Useful in animals &gt; 1 kg &lt; 100 kg</td>
<td></td>
<td>*****</td>
</tr>
<tr>
<td>Useful in animals &lt; 1 kg</td>
<td>*****</td>
<td>*</td>
</tr>
<tr>
<td>Useful in aquatic animals</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Useful in invertebrates</td>
<td>*****</td>
<td>***</td>
</tr>
<tr>
<td>Cost per experiment (animals &gt; 40 kg)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Cost per experiment (animals &lt; 40 kg &gt; 1 kg)</td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>Need for calibration/validation experiments</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Short-term stress to animal</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Long-term stress to animal</td>
<td>***</td>
<td>****</td>
</tr>
<tr>
<td>Data available at fine temporal scales</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Long-term data available</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Accurate data from individuals</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Accurate group data</td>
<td>****</td>
<td>***</td>
</tr>
<tr>
<td>Determination of standard error of estimate</td>
<td>*</td>
<td>****</td>
</tr>
</tbody>
</table>

Acknowledgements

The authors are grateful to Roger Holder for his advice on matters statistical. This article was written while P.J.B., I.L.B. and J.A.G. were supported by a NERC AFI grant.

References


Measuring metabolic rate in the field


Received 24 July 2003; revised 2 September 2003; accepted 5 September 2003

© 2004 British Ecological Society, Functional Ecology, 18, 168–183