

The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls

Kirsten L Rennie and Nicholas J Wareham*

Department of Community Medicine, Institute of Public Health, University of Cambridge, Cambridge CB2 2SR, UK

Submitted 18 June 1998: Accepted 24 August 1998

Abstract

Objective: To review and categorize the problems associated with undertaking physical activity validation studies and to construct a checklist against which any study could be compared.

Results: The studies reviewed demonstrated problems in defining the dimension of physical activity that is of interest and in the selection of an appropriate comparison technique. Ideally this should be closely related to the true exposure of interest and assess that exposure objectively and without correlated error from the study instrument in question. In many studies inappropriate comparison methods have been chosen which do not measure the true underlying exposure and which are likely to have correlated error. The choice of study populations, the frame of reference of the exposure measurement and the use of appropriate statistical methods are also problematic areas.

Conclusions: There is no ideal measurement instrument or validation study design that is suitable for all situations. However, the checklist in this paper provides a means whereby the appropriateness of studies already undertaken or at the planning stage can be assessed.

Keywords
Physical activity
Validation
Questionnaire
Energy expenditure

Physical activity plays an important role in the aetiology of many chronic diseases such as type 2 diabetes, cancer and coronary heart disease¹. However, physical activity is a complex, multidimensional exposure which is difficult to measure, and most of the evidence of its importance has come from studies using subjective questionnaires. Although questionnaires have been sufficient to demonstrate crude associations with disease end-points, uncertainties exist about which dimension of physical activity is being assessed, and the degree to which that assessment is valid².

These uncertainties can lead to difficulties in formulating intervention strategies. For example, as a result of the use of simple questionnaires in studies of the aetiology of type 2 diabetes, it is unclear whether public health interventions should be designed to increase total energy expenditure or to increase fitness². These are distinct public health targets, as it is a different proposition to advocate vigorous activity rather than increasing energy turnover. Therefore, the appropriate design of interventions is dependent on the availability of precise epidemiological data, which is, in turn, dependent upon having valid physical activity measures.

The choice of an appropriate instrument for measuring physical activity in a particular study

depends upon a number of factors, including the main dimension of physical activity that is of health interest, the size of the study and the frame of reference (e.g. current activity or past activity). A wide range of possible field methods exist and the relative merits of each have been discussed in previous reviews^{3–5}. It is not our intention in this article to undertake a similar review, but rather to give a commentary on the notion of validity for these physical activity measurement instruments, to characterize the types of validation studies and the issues around their design, conduct and interpretation. These issues are important not only for researchers planning validation studies, but also for those who plan to apply physical activity measurement instruments in new studies or interpret the results of published studies in which they have been utilized.

Validity and physical activity measurement instruments

In this article, we define validity as the extent to which a measurement instrument assesses the true exposure of interest. This is a different construct to repeatability, which is the extent to which an instrument gives the same result on different occasions. Though this is an important attribute of any instrument, it is not synonymous with validity. Unfortunately, repeatability

*Corresponding author. E-mail njw1004@medschl.cam.ac.uk

is easier to measure and is often given undue attention. Attempts to measure true validity are less frequent.

A particular problem in the physical activity field is that the exposure is multidimensional and some researchers use terms to define the underlying dimensions interchangeably. In the Surgeon General's report on the relationship of physical activity to health¹, physical activity was defined as 'any bodily movement that is produced by the contraction of skeletal muscle and that substantially increases energy expenditure'. Exercise, which is a sub-component of energy expenditure, was defined as 'planned, structured, and repetitive bodily movement done to improve or maintain one or more components of physical fitness'. Physical fitness, 'a set of attributes that people have or achieve that relates to the ability to perform physical activity', includes cardiorespiratory fitness which 'is a health-related component of physical fitness that relates to the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity'. Other dimensions of physical activity include flexibility, weight-bearing and muscular strength.

Clearly an accurate definition of the main exposure of interest is fundamental to the appropriate design of a measurement instrument and its validation. Our focus in this article is on the measurement of physical activity as defined by the Surgeon General's report, which we take to be synonymous with total energy expenditure and not just that component related to programmed exercise.

Validation instruments

A particular problem in validation studies of physical activity is selecting an appropriate comparison method. The ideal validation instrument would objectively measure the true exposure with uncorrelated error from the method being validated⁶. Often it is not possible to use a gold standard, but if a more practical instrument is chosen, it should have a high correlation with the gold standard method⁷. This is particularly true if the epidemiological focus is on an unmeasurable or latent variable, such as usual physical activity, which can only be estimated rather than directly measured⁸.

The definition of the time frame of interest of the questionnaire determines how it should be validated. If the time frame of interest is 1 year, the measurement instrument chosen will be of a different type to that chosen for a short time frame. It might be possible to use a gold standard method as a validation instrument in a short time frame study, but the assessment of longer time frames requires the use of a repeated measures design or an intermediate method which has been validated against the gold standard technique.

If a questionnaire is designed to measure total energy expenditure, then the comparison instrument for a validation study should measure this exposure⁹. Conversely, if the intention is to estimate self-reported leisure time physical activity, then it is appropriate to use other measures of behaviour as the comparison, for example recall diaries. However, in some studies inferences have been made about a physiological construct such as total energy expenditure using a questionnaire that measures self-reported behaviour, when the latter has not been directly compared with the underlying physiological variable.

A wide range of validation instruments have been used in different studies. These vary from other questionnaires through to more objective physiological measures. Given that the focus of this article is on the measurement of the energy expenditure component of physical activity, the following section discusses the appropriateness of different possible validation instruments.

Subjective measures

Some studies have validated questionnaires against other self-reported physical activity measurements. For example, in the US National Health Interview Survey, three subjective questions on job-related, main daily activity and comparison to peers were moderately correlated with a separate continuous energy expenditure score from the questionnaire¹⁰. Recall diaries have also been used as tools to validate physical activity questionnaires. In a study of physical activity in adolescents, for example, a questionnaire of past year physical activity was compared to four 7-day recalls completed over 1 year and reported correlations of 0.55–0.67 for boys and 0.73–0.83 in girls were reported¹¹. Although correlations of this type with other subjective questions or diaries may suggest validity, the validation instruments selected are not of a different type, and may, therefore, be subject to correlated error.

Objective measures

Components of cardiorespiratory fitness have been measured in studies and used as validations for questionnaires. Cardiorespiratory power has been measured by resting heart rate¹², time spent on a maximal graded treadmill exercise test and heart rate response to a submaximal exercise allowing a prediction of maximal VO_2 ^{13–23}. The number of studies that have employed this approach is an indication of the ease of assessing fitness rather than necessarily its epidemiological importance.

The main problem of using a measure of fitness, such as VO_{2max} , as a reference measure is not its accuracy, but its relevance to physical activity or energy

expenditure. Studies of questionnaires concentrating on episodes of vigorous physical activity may find higher correlations with VO_{2max} than those that attempt to measure total daily physical activity^{13,15-17,20-24}, because vigorous activities, which are related to fitness, are more reliably recalled in questionnaires^{25,26}. This does not imply that these questionnaires can be used to measure the totality of physical activity.

Although fitness and total energy expenditure are related, they are not synonymous. Those individuals with high energy expenditure may not necessarily be fit, as measured by VO_{2max} ²². High energy turnover may be the result of prolonged moderate activity, rather than bursts of vigorous activity. In addition, people who spend more time in vigorous activity may compensate by spending less time on moderate activity, and therefore have a relatively low total energy expenditure.

Other validation studies select comparison measures that are related to physical activity, but are not direct measures of it. These include measures of obesity such as body mass index (BMI)^{12,15,20,27}, body composition^{14,20,22}, respiratory function^{14,15,20,22} and high-density lipoprotein (HDL) cholesterol^{12,20,27,28}. The indirectness of these measures might be acceptable if the underlying association with the true exposure were strong. However, in each case this association is only relatively weak^{12,15,27}. Wong *et al.* have demonstrated that if a validation instrument has only a moderate correlation with the underlying true exposure ($\rho_{RT} \leq 0.6$), then it is unlikely to be a good choice⁷. This is particularly true if there is even a moderate degree of correlated error (≥ 0.2), in which case attempts to adjust the observed relationship for measurement error can lead to serious biases⁷.

By contrast, there are other physiological measures that have much higher order of correlation with a gold standard measure of the true exposure. Heart rate monitoring has been used to estimate oxygen consumption and energy expenditure. It has been validated against doubly labelled water and indirect calorimetry with a reported correlation of 0.94²⁹. Subjects wear a heart rate monitor on their chest and minute-by-minute heart rate is recorded. To overcome individual variation in heart rate response to activity, a calibration is used to establish the VO_2 to heart rate relationship for each subject^{30,31}. This technique may be suitable as an objective instrument for validating questionnaires aimed at estimating total energy expenditure, but as yet no such studies have been published.

Caltrac, an accelerometer worn on the hip, has been used in several studies as a validation instrument^{14,22,32-37}. Using a piezometer and the input of the subjects' age, sex, weight and height, an estimate of total energy expenditure and the energy cost of activity

is computed. In theory Caltrac appears to be a good validation instrument since it is an objective, independent measure of physical activity that is unlikely to have correlated error with a physical activity questionnaire. However, in validation studies of Caltrac the degree of correlation is not as high as for heart rate monitoring³⁸. These studies have all been in controlled conditions, and in the field Caltrac is less reliable and more prone to mechanical failure. Nor can it accurately estimate the energy cost of activities such as cycling, and rowing, or weight-bearing movement. Like other movement sensors³⁹, it is not waterproof and cannot be worn during swimming. Subjects are often asked to record these activities separately, but this reduces the objectivity and accuracy of the instrument. This may explain why Caltrac does not correlate with intense activity as expected¹⁴. The observation of a low correlation between Caltrac and questionnaire estimations of physical activity^{14,22,32-36,40} could be a function of the problems of Caltrac as a validation instrument, or may reflect the difficulty of constructing a questionnaire to estimate total energy expenditure. Other movement sensors include the Tritrac system, which records three-dimensional movement and may overcome the problems of estimating energy expenditure of stair climbing and cycling. Vitalog, a solid-state microprocessor with three chest electrodes to determine heart rate and a mercury switch motion sensor attached to the anterior thigh to detect leg and body movement, has been used to compare a 7-day, interviewer-administered activity recall, and correlated well with the activity reported in the recall⁴¹.

A more direct physiological assessment of total energy expenditure can be made using the doubly labelled water method, in which energy expenditure is estimated in free-living individuals by the ingestion of a quantity of water with a known concentration of isotopes of hydrogen and oxygen. From the difference in elimination rates of the two isotopes from the body, carbon dioxide production and oxygen consumption are calculated. Schuit has used the technique to validate the Physical Activity Scale for the Elderly (PASE) questionnaire in 21 elderly people and reported a correlation coefficient of 0.58 (CI = 0.5-0.81)⁴². Similarly, Schulz compared the Modifiable Activity Questionnaire against doubly labelled water⁴³. The technique is expensive both in terms of materials and analysis, thus limiting its application to small validation studies.

Choice of validation study populations

Representativeness

In general the sample population selected for a validation study should reflect the population to whom the questionnaire will be applied. Physical

activity questionnaires are usually designed to measure population-level associations with chronic diseases, such as coronary heart disease or type 2 diabetes. Although these questionnaires are intended for use in general populations including ethnic minority groups, they are often validated in specific subgroups such as relatively active Caucasian populations. The use of university staff and student volunteers, although convenient for recruitment^{14,20,22,34}, introduces the potential for bias as these subgroups may not be representative of the general population. For example, in the Study of Activity, Fitness and Exercise (SAFE) 94% of the subjects were Caucasian, 71% of them held graduate degrees and most were non-smoking²². In addition, a large percentage of this group participated in vigorous activities and had sedentary jobs. In this subgroup, leisure time physical activity forms a major component of total energy expenditure. This may not be true of other occupational groups. Other studies have used groups of young sportsmen⁵⁹ or individuals engaged in an exercise programme⁴¹, both of whom are likely to have higher physical activity levels than are found in the general population.

The problem of participants being self-selected rather than randomly selected is a very difficult one to overcome. As many of the validation methods are relatively intrusive and time-consuming for the participants, self-selection is the only feasible option. Even though self-selection may be inevitable, it is still important to examine the generalizability of the validation subpopulation. Some studies, have poor response rates³⁷ and make no follow-up of non-responders. Therefore, it is not known whether the sample's characteristics and behaviour are representative of the population in which the questionnaire is to be applied.

Exclusions in study populations

The representativeness of the sample for a validation study is affected by the exclusion criteria. Studies using tests such as VO_{2max} as the validation method may face problems of bias as a result of selection on the basis of health, because individuals with angina and previous myocardial infarctions are often excluded from such studies¹⁷. Tests which do not require maximum effort may be preferable in this context because they are more inclusive⁴¹. In the Dutch EPIC cohort, two women were excluded from the analyses as their activity patterns had changed during the study due to retirement³⁷. Such changes in activity patterns would be representative in a subset of the general population and, therefore, exclusion may not be justified. Other studies have excluded people with a BMI > 30¹⁶. But since obese individuals may be less physically active, and may report physical activity differently, the use of a

non-obese validation population limits the groups in which this questionnaire can be used.

Ideally exclusions should be made *a priori* and be as limited as possible to increase the representativeness of the sample to the population in which the questionnaire is to be used.

Gender distribution in the study sample

Many study samples in validation studies have been single sex^{16,18,24,37,41,44,45}. If the questionnaire's intended population is also single sex, then these samples are appropriate. However, results from questionnaires validated in one sex samples may not be generalizable to populations of the opposite or both sexes. For example, in a study of the association between physical activity and diabetes in women, a physical activity questionnaire was said to have been validated 'as a measure of physical activity'⁴⁶. The combined data in the validation study demonstrated a correlation of 0.46 between the self-reported number of days with vigorous activity and VO_{2max} ²¹. However, when the data were stratified by sex, the association was restricted to men and the correlation with women was only 0.26. Therefore, the use of the questionnaire as a validated instrument in an all-female cohort is questionable.

Statistical considerations

Sample size

The sample sizes vary enormously in the validation studies reviewed here, ranging from 17³³ to 986¹⁷. This is largely a function of the validation measure applied. Some instruments, such as other questionnaires, can be administered to relatively large samples. Others (e.g. movement sensors) are only feasible for small samples due to the cost, staff and technical back-up required. The sample size affects the strength of the study's results. In small samples either no significant result is found or large confidence intervals are produced and extreme values may account for much of the observed correlation. Since the validation measurement instrument is often the limiting factor to sample size, the instrument needs to be as simple and cost-effective as possible without compromising accuracy in order for strong correlations to be found.

Good descriptions of how to calculate the sample size for different forms of study have been produced for nutritional epidemiology⁴⁷ and the issues in physical activity are analogous. For example, if the intention is to estimate population level energy expenditure using heart rate monitoring, the number of subjects required can be computed with knowledge of the standard deviation from previous studies³¹. To calculate mean population level PAL (physical activity level) within 0.15 with a significance of 5% at

90% power, requires⁴⁷

$$n = 2\sigma^2(Z_{\alpha/2} + Z_{\beta})^2/d^{*2}.$$

If we assume that the standard deviation is 0.31³¹, then

$$n = 2(0.31)^2 \times (10.5)/0.0225 = 90.$$

How many days of measurement are necessary?

If the goal of a study is to measure current energy expenditure, it is possible, using standard equations, to compute the number of days of testing that are required to characterize current energy expenditure⁶. A reanalysis of data in a previous study³⁰, shows that the correlation between energy expenditure on 2 adjacent days is 0.65. Therefore, measuring for 2 days would yield a new exposure measurement, with a validity coefficient of 0.79, 3 days would give 0.85 and 4 days would give 0.88. In reality the choice of the number of days is usually dependent upon pragmatic considerations such as the cost and the willingness of the subject to be monitored. If the goal of the study is not the measurement of current activity, but rather usual activity, a different approach needs to be taken. Usual activity is a latent variable which is not directly measurable. It can, however, be estimated by taking repeated measurements throughout the course of the frame of reference, for example 1 year. This form of approach allows the relationship between the underlying latent variable and the measurement of outcome to be estimated³¹ and is an approach that has been previously used in nutritional epidemiology⁴⁸.

Problems of analysis

As with the computation of sample size, direct analogies for analysing validation studies of physical activity can be taken from the nutritional epidemiological field. The approach to analysis is dependent on the type of outcome measured and whether it is reduced to a continuous ordered categorical or discreet categorical variable. All too often questionnaires, which are said to have been validated, are undermined by validation studies which have actually used inappropriate statistical methods. For example, the validation study by Siconolfi, which is used as justification for the Nurse's Health Study, used linear regression to analyse the relationship between the number of sweating episodes per week with maximal oxygen uptake²¹. This is an inappropriate method since regression analysis is designed for studying the relationship between two continuous variables. This study analysis may also have been flawed as the data were skewed with a few outliers biasing the extent of the correlation.

Due to the ranges of the category levels adopted for

physical activity assessment, often the distributions of these levels are skewed towards the low-activity end. This means that the data are often not suitable for straightforward parametric statistical analysis and other methods need to be employed.

Another type of problem occurs when single-item self-assessment questions have been compared to composite scores derived from an entire questionnaire^{10,12}. If the single question is part of the calculation of the index or composite score, then problems of autocorrelation may result¹³.

Conclusions

In this article we have grouped together some of the problems that face researchers attempting to undertake or seeking to interpret validation studies of physical activity measurement instruments. None of these criticisms implies that there are easy answers, but the alternative to debating and categorizing these problems is to ignore them and pretend that they do not exist. The epidemiological knowledge about the importance of an exposure, such as physical activity, passes through a process of evolution. When little is known, simple measures of exposure can be used to demonstrate disease associations. Simplicity and inexpense are key requisites because few funders will commit resources to investigate speculative exposures in a more detailed fashion. Epidemiologists can always claim that because of the principle of attenuation, simple exposure measurement will always result in an underestimation of the true exposure-disease association. However, this assumption assumes that the simplified measurement of the exposure is unbiased and that the same underlying true exposure is being assessed. In the case of physical activity this may not necessarily be true. Once understanding has moved on from the description of simple relationships, a different approach has to be taken to the measurement of exposure. Interest then moves on from simple questions about whether physical activity is related to disease, to more involved questions concerning which aspects of activity are important and quantitative questions about how much activity is necessary. The measurement instruments that are required to answer these questions must be more accurate. As physical activity epidemiology is beginning to enter this phase of its evolution, it is timely to consider whether the measurement instruments that are currently available meet the standards necessary for this more exacting role.

We have not sought to propose an ideal measurement instrument because the choice for any given study depends upon a wide range of factors. However, more careful attention perhaps ought to be given to the notion of validity than has sometimes been the case.

Table 1 Checklist for the validation of physical activity instruments

-
- 1 Has the dimension of physical activity that the instrument is purported to measure been clearly defined?
 - 2 Does the validation method chosen measure the true exposure of interest and has it been applied in the same time frame of reference?
 - 3 Has correlated error between the validation method and the physical activity instrument been avoided as far as possible?
 - 4 Is there a close relationship between the validation method and the appropriate 'gold standard' instrument?
 - 5 Is the sample chosen representative of the population to whom the physical activity instrument will be administered?
 - 6 Have appropriate statistical techniques been employed to assess the validity of the physical activity instrument?
-

The claim that a particular method has been 'validated' ought really to be questioned by anyone proposing to use that method. The checklist summarized in Table 1 provides a means by which researchers can gauge the likelihood that a validation study has achieved its goal.

Acknowledgements

NJ Wareham is an MRC clinician scientist fellow. KL Rennie holds an MRC studentship.

References

- 1 US Department of Health and Human Services. *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.
- 2 Wareham N, Rennie K. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Int. J. Obesity* 1998; **22**(S2): S30–8.
- 3 LaPorte RE, Montoye HJ, Caspersen CJ. Assessment of physical activity in epidemiologic research: problems and prospects. *Publ. Health Rep.* 1985; **100**: 131–46.
- 4 Drury TF ed. *National Center for Health Statistics. Assessing Physical Fitness and Physical Activity in Population-Based Surveys*. Public Health Service, Washington: US Government Printing Office, 1989.
- 5 Kriska AM, Caspersen CJ (eds) A collection of physical activity questionnaires for health-related research. *Med. Sci. Sports Exerc.* 1997; **29**(6 Suppl): S1–S205.
- 6 Armstrong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology*. Oxford: Oxford University Press, 1994.
- 7 Wong MY, Day NE, Wareham NJ. The design of validation studies II: the multivariate situation. *Stat. Med.* 1998; in press.
- 8 McCutcheon AL. Latent class analysis. In: Drury TF, ed. *Assessing Physical Fitness and Physical Activity in Population-based Surveys*. Hyattsville, Maryland: National Center for Health Statistics, 1989: 633–49.
- 9 Caspersen CJ. Physical activity epidemiology: concepts, methods, and applications to exercise science. *Exerc. Sports Sci. Rev.* 1989; **17**: 423–73.
- 10 Weiss TW, Slater CH, Green LW, Kennedy VC, Albright DL, Wun C. The validity of single-item, self assessment questions as measures of adult physical activity. *J. Clin. Epidemiol.* 1990; **43**: 1123–9.
- 11 Aaron DJ, Kriska AM, Dearwater SR, Cauley JA, Metz KF, LaPorte RE. Reproducibility and validity of an epidemiologic questionnaire to assess past year physical activity in adolescents. *Am. J. Epidemiol.* 1995; **142**: 191–201.
- 12 Mayer EJ, Alderman BW, Regensteiner JG *et al.* Physical activity assessment measures compared in a biethnic population: the San Luis Valley diabetes study. *Am. J. Clin. Nutr.* 1991; **53**: 812–20.
- 13 Singh PN, Tonstad S, Abbey DE, Fraser GE. Validity of selected physical activity questions in white seventh-day adventists and non-adventists. *Med. Sci. Sports Exerc.* 1996; **28**: 1026–37.
- 14 Ainsworth BE, Jacobs DR, Leon AS, Richardson MT, Montoye HJ. Assessment of the accuracy of physical activity questionnaire occupational data. *JOM* 1993; **35**: 1017–27.
- 15 Hopkins WG, Wilson NC, Russell DG. Validation of the physical activity instrument for the life in New Zealand survey. *Am. J. Epidemiol.* 1991; **133**: 73–82.
- 16 Elosua R, Marrugat J, Molina L, Pons S, Pujol E. Validation of the Minnesota leisure time physical activity questionnaire in Spanish men. *Am. J. Epidemiol.* 1994; **139**: 1197–209.
- 17 Booth ML, Owen N, Bauman A, Gore CJ. Relationship between a 14-day recall measure of leisure-time physical activity and a submaximal test of physical work capacity in a population sample of Australian adults. *RQES* 1996; **67**: 221–7.
- 18 Knapik J, Zoltick J, Rottner CH *et al.* Relationships between self-reported physical activity and physical fitness in active men. *Am. J. Prev. Med.* 1993; **9**: 203–8.
- 19 Sulieman S, Nelson M. Validation in London of a physical activity questionnaire for use in a study of postmenopausal osteopaenia. *J. Epidemiol. Comm. Health* 1997; **51**: 365–72.
- 20 Lamb KL, Brodie DA. Leisure-time physical activity as an estimate of physical fitness: a validation study. *J. Clin. Epidemiol.* 1991; **44**: 41–52.
- 21 Siconolfi SF, Lasater TM, Snow RCK, Carleton RA. Self-reported physical activity compared with maximal oxygen uptake. *Am. J. Epidemiol.* 1985; **122**: 101–5.
- 22 Jacobs DR, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med. Sci. Sports Exerc.* 1993; **25**: 81–91.
- 23 Berthouze SE, Minaire PM, Castells J, Busso T, Vico L, Lacour J. Relationship between mean habitual daily energy expenditure and maximal oxygen uptake. *Med. Sci. Sports Exerc.* 1995; **27**: 1170–9.
- 24 Kohl HW, Blair SN, Paffenbarger RS, Macera CA, Kronenfeld JJ. A mail survey of physical activity habits as related to measured physical fitness. *Am. J. Epidemiol.* 1988; **127**: 1228–39.
- 25 Slattery ML, Jacobs DR. Assessment of ability to recall physical activity of several years ago. *AEP* 1995; **5**: 292–6.
- 26 Blair SN, Kannel WB, Kohl HW, Goodyear N, Wilson PWF. Surrogate measures of physical activity and physical fitness. *Am. J. Epidemiol.* 1989; **129**: 1145–55.

- 27 Washburn RA, Smith KW, Goldfield SRW, McKinlay JB. Reliability and physiologic correlates of the Harvard alumni activity survey in a general population. *J. Clin. Epidemiol.* 1991; **44**: 1319–26.
- 28 Washburn RA, Goldfield SRW, Smith KW, McKinlay JB. The validity of self-reported exercise-induced sweating as a measure of physical activity. *Am. J. Epidemiol.* 1990; **132**: 107–13.
- 29 Ceesay SM, Prentice AM, Day KC, Murgatroyd PR, Goldberg GR, Scott W. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *Br. J. Nutr.* 1989; **61**: 175–86.
- 30 Wareham NJ, Hennings SJ, Prentice AM, Day NE. Feasibility of heart rate monitoring to estimate total level and pattern of energy expenditure in a population-based epidemiological study. *Br. J. Nutr.* 1997; **78**: 889–900.
- 31 Wareham NJ, Hennings SJ, Byrne CD, Hales CN, Prentice AM, Day NE. A quantitative analysis of the relationship between habitual energy expenditure, fitness and the metabolic cardiovascular syndrome. *Br. J. Nutr.* 1998; in press.
- 32 Williams E, Klesges RC, Hanson CL, Eck LH. A prospective study of the reliability and convergent validity of three physical activity measures in a field research trial. *J. Clin. Epidemiol.* 1989; **42**: 1161–70.
- 33 Kriska AM, Knowler WC, LaPorte RE *et al.* Development of questionnaire to examine relationship of physical activity and diabetes in Pima Indians. *Diabetes Care* 1990; **13**: 401–11.
- 34 Richardson MT, Ainsworth BE, Wu H-C, Jacobs DR, Leon AS. Ability of the atherosclerosis risk in communities (ARIC)/Baeke questionnaire to assess leisure-time physical activity. *Int. J. Epidemiol.* 1995; **24**: 685–93.
- 35 Richardson MT, Leon AS, Jacobs DR, Ainsworth BE, Serfass R. Comprehensive evaluation of the Minnesota leisure time physical activity questionnaire. *J. Clin. Epidemiol.* 1994; **47**: 271–81.
- 36 Miller DJ, Freedson PS, Kline GM. Comparison of activity levels using the Caltrac accelerometer and five questionnaires. *Med. Sci. Sports Exerc.* 1994; **26**: 376–382.
- 37 Pols MA, Peeters PHM, Kemper HCG, Collette HJA. Repeatability and relative validity of two physical activity questionnaires in elderly women. *Med. Sci. Sports Exerc.* 1996; **28**: 1020–5.
- 38 Montoye HJ, Kemper HCG, Saris WHM, Washburn RA. *Measuring Physical Activity and Energy Expenditure.* Human Kinetics, 1996: 79–89.
- 39 Sequeira MM, Rickenbach M, Wietlisbach V, Tullen B, Schutz Y. Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. *Am. J. Epidemiol.* 1995; **142**: 989–99.
- 40 Pols MA, Peeters PHM, Bueno-de-Mesquita BH *et al.* Validity and repeatability of a modified Baecke questionnaire on physical activity. *Int. J. Epidemiol.* 1995; **24**: 381–8.
- 41 Taylor CB, Coffey T, Berra K, Iaffaldano R, Casey K, Haskell WL. Seven-day activity and self-report compared to a direct measure of physical activity. *Am. J. Epidemiol.* 1984; **120**: 818–24.
- 42 Schuit AJ, Schouten EG, Westerup KR, Saris WHM. Validity of the Physical Activity Scale for the Elderly (PASE): according to energy expenditure assessed by the doubly labeled water method. *J. Clin. Epidemiol.* 1997; **50**: 541–6.
- 43 Schulz LO, Harper IT, Smith CJ, Kriska AM, Ravussin E. Energy intake and physical activity in Pima Indians: comparison with energy expenditure measured by doubly-labeled water. *Obesity Res.* 1994; **2**: 541–8.
- 44 Wolf AM, Hunter DJ, Colditz GA *et al.* Reproducibility and validity of a self-administered physical activity questionnaire. *Int. J. Epidemiol.* 1994; **23**: 991–9.
- 45 Albanes D, Conway JM, Taylor PR, Moe PW, Judd J. Validation and comparison of eight physical activity questionnaires. *Epidemiology* 1990; **1**: 65–71.
- 46 Manson JE, Rimm EB, Stampfer MJ *et al.* Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet* 1991; **338**: 774–8.
- 47 Cole TJ. Sampling, study size, and power. In: Margetts BM, Nelson M, eds. *Design Concepts in Nutritional Epidemiology.* Oxford: Oxford University Press, 1991: 53–78.
- 48 Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology.* Oxford: Oxford University Press, 1991.