Chapter 1

Characteristics of weight control and physical activity

Weight control

Weight control is widely defined as approaches to maintaining weight within the 'healthy' (i.e., 'normal' or 'acceptable') range of body mass index of 18.5 to 24.9 kg/m² throughout adulthood (WHO Expert Committee, 1995). It should also include prevention of weight gain of more than 5 kg in all people. In those who are already overweight, a reduction of 5–10% of body weight is recommended as an initial goal.

Anthropometric measures
Body mass index

When we speak about the prevalence of obesity in populations, we mean the fraction of people who have excess storage of body fat. In adult men with weight in the acceptable range, the percentage of body fat is around 15–20%. In women, this percentage is higher (about 25–30%). Because differences in weight between individuals are only partly due to variations in body fat, many people object to the use of weight or indices based on height and weight (such as the body mass index) to discriminate between overweight and normal-weight people. Body mass index (BMI) is a measure of body mass relative to height, calculated as weight (kg) divided by height squared (m²). Examples can of course be found to illustrate inappropriate use of BMI to compare certain individuals, such as an identical body mass index in a young male body builder and a middle-aged obese woman. In general, however, there is a very good correlation between BMI and the percentage of body fat in large populations. Deurenberg et al. (1991) showed that, in Dutch adults, the following equation can be used to estimate the body fat percentage:

$$\text{Percentage body fat} = 1.2 \times \text{BMI} + 0.23 \times (\text{age}) - 10.8 \times (\text{gender}) - 5.4$$

In this equation the value for gender is one for men and zero for women.

About 80% of the variation in body fat between individuals can be explained by this formula, with a standard error of about 4%. It follows from the equation that for a given height and weight, the body fat percentage is about 10% higher in women than in men. Also, people get fatter when they get older even when their body weight is stable. The good correlation between BMI and fat percentage implies that, at the population level, BMI can be used to classify people in terms of excess body fat. In practice, people or populations are usually not classified on the basis of their body fat percentage but of their BMI. Usually, the same cut-points are applied for men and women and for different age-groups, because the relationships between BMI and mortality are similar (i.e., the relative mortality associated with obesity is similar in men and women). However, in most age-groups, the absolute mortality is much lower in women, implying that the effect of excess body fat is less in women than BMI in men. This may be because in women the excess body fat is usually distributed as subcutaneous fat and mainly peripherally (thighs, buttocks, breasts), while in men there is a relative excess of body fat stored in the abdominal cavity and as abdominal subcutaneous fat. It has been suggested that the optimal body mass index (i.e., the BMI associated with lowest relative risk) increases with age (Andres, 1985).

Fat distribution patterns

Fat can be stored in adipose tissue as subcutaneous fat and as intra-abdominal fat. The pattern of subcutaneous fat can differ greatly with age, sex and ethnicity. Women tend to store subcutaneous fat in the gluteal and femoral regions and breasts, whereas men tend to store subcutaneous fat more in the truncal region. Intra-abdominal (or visceral) fat is formed by fat deposition in the omentum and mesentery and as retro-peritoneal fat. The omental and mesenteric adipose tissues are drained by the portal vein and are sometimes labelled as 'portal fat' (Björntorp, 1990). Apart from their unique location and venous drainage, the portal tissues have several other characteristics that make them liable to involvement in the metabolic disturbances associated with obesity. The fat cells here are more responsive to lipolytic stimuli (such as epinephrine and norepinephrine) and less responsive to the anti-lipolytic effect of insulin. The result can be overproduction of free fatty acids which are released into the portal vein and thus expose the liver to relatively high concentrations of free fatty acids, implicated in the development of heart disease and diabetes (Björntorp, 1990).

Since the pioneering work of Jean Vague in the 1940s, it has slowly become accepted that different body morphology or types of fat distribution are independently related to the health risks associated with obesity (Vague, 1956). Starting with Vague's brachio-femoral adipo-muscular ratio as an index of fat distribution (which was based on ratios of skinfolds and circumferences of the arms and thighs), more recent indices were designed specifically to be good predictors of intra-abdominal fat. The most popular is the waist to hip circumference ratio (WHR). The simplest measure is the waist circumference, which may be predictive of intra-abdominal fat at least as accurately as the WHR (Pouliot et al., 1994) and levels of cardiovascular risk factors and disease as well as BMI and WHR (Han et al., 1995). It has been suggested that waist circumference could be used to replace classifications based on BMI and the WHR (Lean et al., 1995; Booth et al., 2000) and this has been agreed by the WHO. More complex measures, such as the sagittal abdominal diameter, the ratio of waist/thigh circumference, the ratio of waist/height or the conicity index, have also been proposed to perform even better than waist circumference for one or more of these purposes. However, the differences between these measures are small and the use of ratios may complicate the interpretation of associations with disease and their consequences for public health. For instance, the waist/height ratio may be a better predictor of morbidity because the waist circumference is positively associated with disease, and because height, for reasons unrelated to body composition or fat distribution, is inversely associated with cardiovascular disease risk.

Replacing BMI and WHR by simple cut-points which are optimal for each sex, age-group, population and relationship with specific diseases may, however, be too simple. Still, as suggested by Lean et al. (1995), some cut-points may provide guidance in interpreting values of waist circumference for adults (Table 1). Other cut-points, based on classification of subjects on a 'critical level' of intra-abdominal fat, have been proposed (Lemieux et al., 1996).

Definition of obesity, overweight and underweight

Cut-points of BMI, that apply to both men and women and to all adult age-groups, have been proposed by a WHO Expert Committee for the classification of overweight (WHO Expert Committee, 1995) (Table 2). The BMI cut-points for degrees of under- and over-weight are largely arbitrary. The cut-points for overweight (25, 30, 40 kg/m^2) were initially based on the monotonic increase in the risk of mortality throughout the range of 20 to 40 kg/m^2. The cut-point for under-weight (18.5 kg/m^2) was largely based on health-related problems associated with malnutrition in developing countries.

There are limitations in the interpretation of BMI in very old subjects as well as in certain ethnic groups with unusual body proportions (e.g., populations where stunted growth is common or those with relatively short leg length compared to sitting height).

Causes of obesity

Obesity is always caused by an excess of energy intake over energy expenditure, referred to as a positive energy balance. Factors influencing this are biological (e.g., age, sex, genes), environmental and behavioural (including diet and physical activity). It has been argued that the prevalence of obesity in populations is determined mainly by the physical, economical and socio-cultural environment (Egger & Swinburn, 1997), which may act at a macro-level (i.e., on nations or large populations) or at a micro-level (a local or household level). These factors determine what percentage of the population is or will become

Table 1. Sex-specific cut-points for waist circumference

	Level 1[a]	Level 2[b]
Men	≥ 94 cm (~ 37 inches)	≥ 102 cm (~ 40 inches)
Women	≥ 80 cm (~32 inches)	≥ 88 cm (~ 35 inches)

[a] Level 1 was initially based on replacing the classification of overweight (BMI ≥ 25 kg/m^2) in combination with high WHR (≥ 0.95 in men and ≥ 0.80 in women)

[b] Level 2 was based on classification of obesity (BMI ≥ 30 kg/m^2) in combination with high WHR (Han et al., 1995; Lean et al., 1995).

Table 2. Cut-points of body mass index for the classification of weight

BMI	WHO classification	Popular description
< 18.5 kg/m^2	Underweight	Thin
18.5–24.9 kg/m^2	–	'Healthy', 'normal' or 'acceptable' weight
25.0–29.9 kg/m^2	Grade 1 overweight	Overweight
30.0–39.9 kg/m^2	Grade 2 overweight	Obesity
≥ 40.0 kg/m^2	Grade 3 overweight	Morbid obesity

Source : WHO Expert Committee, 1995

obese, but do not explain which individuals are likely to become obese; at the individual level, the biological and behavioural factors are the primary determinants of obesity.

Diminished physical activity, changing dietary patterns, energy-dense diets and inadequate adjustment of energy intake relative to diminished energy requirements are all likely to be major determinants of the observed changes in the prevalence of obesity over time. Prentice and Jebb (1995) proposed that, at the population level, limited physical activity was more important than energy or fat consumption in explaining the time trends of obesity in the United Kingdom. Their analysis was based on aspects of physical activity (such as number of hours spent watching television) and household consumption survey data. Although such ecological correlations seem compelling, they may also be misleading. One could, for example, find very impressive correlations over time when plotting the density of mobile

phones versus the prevalence of obesity, without any likely mechanism.

The influence of dietary intake on the prevalence of obesity in populations and individuals is a very difficult subject to study, as it is usually necessary to rely on self-reporting of diet. In particular, energy and fat consumption are known to be underreported with increasing degrees of overweight (Seidell, 1998).

Changes in smoking behaviour may also contribute to changes in body weight at a population level. Data from the United States show that, although smoking cessation can explain some of the increase in the prevalence of overweight, it cannot on its own account for the major portion of the increase (Flegal et al., 1995). Other studies have suggested that the increase in prevalence of obesity may be independent of smoking status (Boyle et al., 1994; Wolk & Rössner, 1995).

Epidemiological methods based on self-reporting, used to assess energy intake and energy expenditure, are not

only subject to bias but also have a high ratio of within- to between-subject variation. Even subtle disruptions in energy balance can explain weight gain over time in individuals or populations (Figure 3). In populations, an increase of one unit in BMI corresponds to an increase of about 5% in the prevalence of obesity. Assuming height remains constant, an average weight increase of slightly less than 3 kg corresponds to an increase in one unit of BMI. If this occurs over a ten-year period, the excess number of calories ingested need only be of the order of 87 000 kJ (21 000 kcal) over ten years (i.e. about 21 kJ (5 kcal) per day). In this theoretical calculation, we ignore the effect of the increased energy expenditure that results from a weight increase (about 190–250 kJ (45–60 kcal) per day for a 3-kg weight increase). Nevertheless, an energy imbalance of about 210 kJ (50 kcal) per day is easily achieved. It is clear that such small persistent changes in energy balance over several years are not detectable

(a) A fine balance: 600 kJ (140 kcal) excess intake is equivalent to:

one croissant

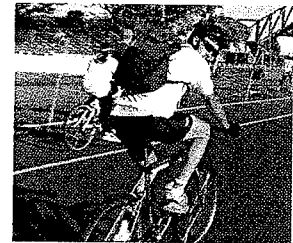one piece of chocolate cake

one bottle of beer

handful of peanuts

(b) Energy expenditure of 600 kJ

= 14 minutes

= 19 minutes

= 21 minutes

= 35 minutes

Figure 3 A fine balance: examples of (a) intake and (b) expenditure of 600 kJ (140 kcal)

by existing methods for measuring energy expenditure and energy intake in populations (WHO Expert Committee, 1995).

At the population level, some other characteristics are associated with the prevalence of obesity, which in individuals is the result of a long-term positive energy balance. These include:

♦ Age: obesity increases at least up to age 50 to 60 years in men and women
♦ Gender: women generally have a higher prevalence of obesity compared with men, especially above 50 years of age
♦ Ethnicity: there are large, usually unexplained, variations between ethnic groups
♦ Educational level and income: in industrialized countries, prevalence of obesity is higher in those with lower education and/or income
♦ Marital status: obesity tends to increase after marriage
♦ Parity: BMI may increase with increasing number of children. Although this contribution seems to be, on average, less than 2 kg per pregnancy (Williamson et al., 1994), pregnancy-related weight gain may be a significant contribution to weight gain for some women, especially those not lactating.
♦ Smoking: smoking lowers body weight and cessation of smoking leads to an increase. The associations between smoking and obesity may, however, vary considerably between populations (Molarius et al., 1997).
♦ Alcohol consumption: the effect is unclear in most populations

Obesity also has a strong familial component. By combining evidence from twin studies, adoption studies, family studies and other relevant data, it has been calculated that the heritability of overweight is in the range 25–40%

(Bouchard, 1994) (Figure 4). Studies in animal models of genetic obesity have revealed important pathways of energy homeostasis, e.g., the role of leptin (Friedman & Halaas, 1998), and gene mutations have been identified that are rare causes of human obesity. For common human obesity, however, the evidence for genetic factors is still fragmentary and incomplete (Chagnon et al., 2000). It is very likely that obesity is a multifactorial polygenic trait. Gene–environment and gene–gene interactions are likely to play an important role.

Timing of obesity

Although it is clear that obesity can develop at any age, there are several critical periods when humans seem to be more liable to accumulate body fat (Dietz, 1994).

♦ Prenatal growth: there is evidence that low birth weight is associated with abdominal fatness in middle-age (Dietz, 1994)
♦ Adiposity rebound period: when BMI is plotted against age, there is first a sharp reduction of BMI from birth until about six years of age. There is then a levelling off followed by an increase of BMI with age. Dietz (1994) has proposed that an early adiposity rebound is predictive of obesity in adults.
♦ Adolescence: physical activity in affluent societies declines rapidly with age in adolescents, the decline being especially pronounced between 15 and 18 years (Caspersen et al., 2000).
♦ Young adulthood: a period of rapid weight gain is often observed in young adults aged 25–40 years. This is usually a period of great behavioural change. People become settled, have children, buy a car and work long hours to build a career.
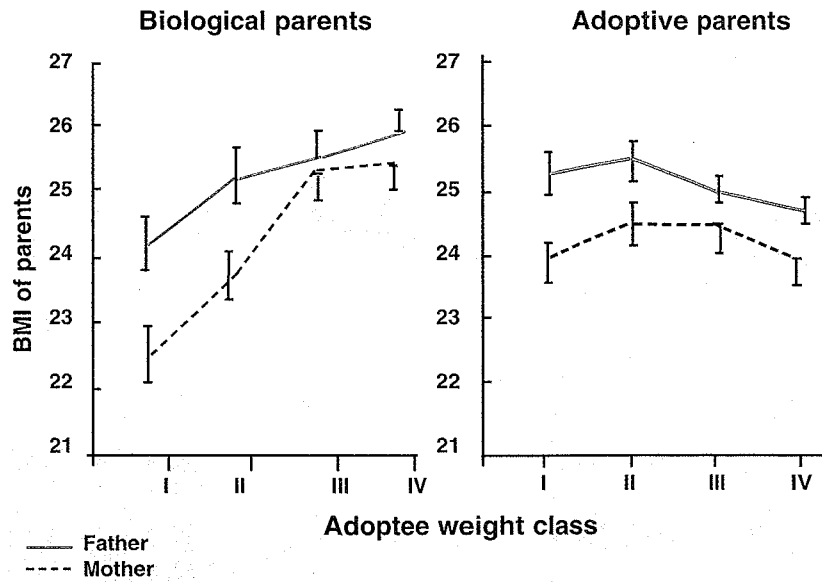
Figure 4 Genes versus environment: the BMI of adults who were adopted as children is clearly correlated with that of their biological parents but not that of the adoptive parents (I, thin; II, median, III, overweight; IV, obese)
Adapted from Stunkard et al. (1986)

Such changes all may reduce physical activity levels.

♦ Menopause: BMI tends to level off in men by the age of 50–60 years but continues to increase in women. Women tend to continue to gain weight after menopause and to accumulate more fat as abdominal fat compared to their premenopausal period.

Measures of weight (critical assessment of the various measures used)

BMI, WHR and waist circumference (Figure 5) are the most common measures used to estimate overweight, obesity and relative body composition in epidemiological studies. The only widely accepted criteria for obesity are based on BMI. New methods based on electrical resistance and impedance, magnetic resonance imaging and computer-assisted tomography are also used, but are expensive and seldom applicable in epidemiological studies. This chapter therefore concentrates on simple anthropometric methods.

The reproducibility and validity of weight and height measurements are high (Willett, 1998). In general, body weight is among the most precise biological measurements, even under imperfect conditions. However, in many epidemiological studies, weight and height are based on self-reports and it is known that people tend to underreport their weight and overreport their height slightly (Figure 6). As a consequence, BMI based on self-reported data will be biased downward. The degree of underreporting is proportional to the degree of overweight, age, and socioeconomic status (Niedhammer et al., 2000). Epidemiological measures of association, such as relative risks, however, are not appreciably affected by this degree of measurement error. In contrast, comparisons of obesity prevalence between populations will be invalid if some data are based on actually measured weight and height while others are based on

self-reported values. For instance, a study in Australia showed that 62% of men and 47% of women were classified as overweight or obese based on measured height and weight compared with 39% and 32%, respectively, based on self-reported height and weight (Flood et al., 2000). Even recalled weight from many years earlier is highly valid, although the error is greater than for self-reported current weight. Correlations between measured weight at 18–30 years of age and recalled weight 20–30 years later usually are about 0.80 (Rhoads & Kagan, 1983; Stevens et al., 1990; Must et al., 1993; Troy et al., 1995).

The validity of BMI as a measure of obesity is generally high. In studies where the reference method has been underwater weighing, the correlation between BMI and densitometry-estimated body fat has generally been 0.60–0.70 in adults (see Willett, 1998). Although BMI is primarily thought of as an estimate of percentage body fat, it is more correctly a measure of absolute fat mass adjusted for height. Thus, considerably higher correlations (0.82–0.91) between BMI and absolute fat mass adjusted for height have been found (Spiegelman et al., 1992).

Even though BMI is an excellent measure of adiposity in young and middle-aged adults, it is less useful in older adults. Many elderly people lose lean body mass, so that for the same BMI, the percentage of fat mass increases (Gallagher et al., 1996). Thus other measures of adiposity may be more appropriate for the elderly (Willett, 1998). For example, changes in abdominal circumference reflect adipose rather than muscle tissue and may thus be a better indicator of overall adiposity than weight alone or BMI.

There are also differences between populations in body build. For example, the relationship between percentage body fat and BMI is different between Singaporeans and Caucasians and also
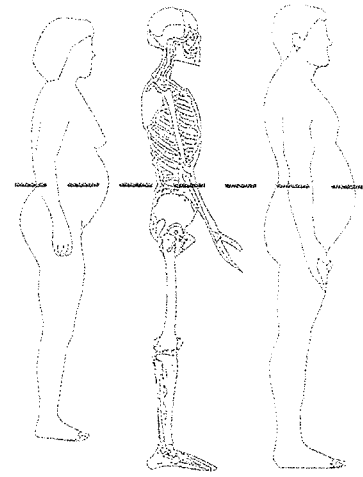
Figure 5 Waist circumference measurement. To measure waist circumference, locate the upper hip bone and the top of the right iliac crest. Place a measuring tape in a horizontal plane around the abdomen at the level of the iliac crest. Before reading the tape measure, ensure that the tape is snug, but does not compress the skin, and is parellel to the floor. The measurement is made at the end of a normal expiration.

Figure 6 Studies have demonstrated that obese people have a tendency to underreport their weight

between the Chinese, Malaysians and Indians in Singapore (Deurenberg-Yap *et al.*, 2000). For the same amount of body fat as for Caucasians who have BMI 30 kg/m² or over, the cut-points for obesity would have to be about 27 for Chinese and Malays and 26 for Indians. In contrast, Polynesians living in New Zealand (Maoris and Samoans) have a significantly higher ratio of lean body mass to fat mass compared with Europeans (Swinburn *et al.*, 1999). Thus, at higher BMI levels, Polynesians are significantly leaner than Europeans, suggesting the need for ethnic-specific BMI definitions of overweight and obesity.

With the increasing evidence on the health risks associated with a predominance of abdominal (visceral) fat, numerous anthropometric indicators of abdominal obesity have been proposed (Molarius & Seidell, 1998). Of these indicators, waist circumference and WHR are widely used in epidemiological studies as well as in screening persons at risk of chronic diseases.

The WHR is difficult to measure in a strictly standardized way and also difficult to interpret biologically. The waist circumference predominantly measures visceral organs and abdominal (both subcutaneous and intra-abdominal) fat. The hip circumference may reflect various aspects of body composition, such as muscle mass, fat mass and skeletal frame. When these two circumferences are combined in a ratio, it is difficult to interpret differences between or within individuals (Molarius & Seidell, 1998). In addition, a reduction in weight usually results in a reduction in both waist and hip circumference, so will not necessarily result in a change in WHR.

Waist circumference is strongly correlated with visceral fat areas and can easily be measured and interpreted. This makes it a suitable candidate for an indicator of abdominal obesity. It is also well correlated with overall fatness.

Waist circumference is strongly correlated with BMI and adding waist measurements to age and BMI does not much improve the explanation of the variance in visceral fat, especially in women (Seidell *et al.*, 1988). Waist circumference is also strongly correlated with abdominal subcutaneous fat, total abdominal fat and total body fat (Lean *et al.*, 1996).

Cut-points have been suggested for both waist circumference and WHR (Molarius & Seidell, 1998), based on results obtained in Caucasian populations. A detailed analysis of 19 populations in the WHO MONICA study showed that the optimal screening cut-points for waist circumference may be population-specific (Molarius *et al.*, 1999). At waist action level 2 (waist circumference 102 cm or more in men and 88 cm or more in women, respectively; BMI 30 kg/m² or more), sensitivity varied from 22% to 64% in men and from 26% to 67% in women, whereas specificity was >95% in all populations. Sensitivity was in general lowest in populations in which overweight was relatively uncommon and highest in populations with a relatively high prevalence of overweight.

Even though criteria for waist circumference and WHR to be used in the public health setting are difficult to determine, these measurements should be applied in epidemiological studies, as the use of different measures of adiposity may give further insight into the etiology of disease.

Physical activity
Definition of physical activity
Physical activity is defined as bodily movement that is produced by the contraction of skeletal muscle and that substantially increases energy expenditure (US Department of Health and Human Services, 1996) (Figure 7). It has three main components:
♦ Occupational work
♦ Household, garden and other domestic activities

♦ Leisure-time physical activity (including exercise and sport)

By the terms physical inactivity or sedentary behaviour, we mean "a state in which body movement is minimal and energy expenditure approximates the resting metabolic rate".

Measurement of physical activity
To determine the relationship between physical activity and cancer, it is necessary to obtain valid and reproducible measurements of exposure and outcome variables. This can be difficult, since physical activity is a very complex behaviour that can be measured in many ways. Further, methods for measuring human energy expenditure are either precise but very restrictive – and thus limited to use over a short period of time – or are less restrictive and usable over longer periods but of lower precision.

Available methods to estimate physical activity and total energy expenditure include:
♦ Calorimetry (direct and indirect)
♦ Physiological markers (e.g., using doubly labelled water)
♦ Mechanical and electronic monitors (e.g., of heart rate, pedometers)
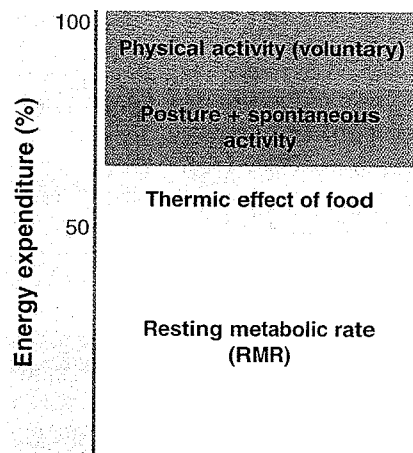♦ Behavioural observation

Figure 7 Components of daily energy expenditure
Adapted from: US DHHS, 1996

+ Job classification
+ Surveys (indirect calorimetric diaries, recall questionnaires, quantitative histories, and lifetime individual histories)

The operational definition of physical activity will vary not only according to the measurement method used, but also according to the type of research being done.

The biological and physiological approaches to assessment of physical activity, i.e., calorimetry, physiological and mechanical monitors, are not applicable in large population studies, but provide objective data for validating population surveys. The method using doubly labelled water has potential to bridge the gap. This involves measuring integral production of carbon dioxide for up to three weeks by the difference in elimination rates of the stable isotopes deuterium and oxygen-18 from doubly labelled water after ingestion of a quantity of water enriched with both isotopes. At present, however, such labelled water is expensive and not widely available.

The most convenient and commonly used measures of physical activity in cancer epidemiology have been job classification and surveys. Questionnaires and interviews used in such surveys are usually unrestricted but imprecise. Because of the low costs involved, they remain the most frequently used methods for assessment of physical activity, especially in epidemiological studies. The validity and test–retest reproducibility of questionnaires and/or interviews concerning physical activity have not been extensively studied. In general, strenuous physical activity appears to be recalled with greater accuracy than light or moderate-intensity activity, whether the recall is for recent periods or earlier periods of time (Slattery & Jacobs, 1995). It is useful to include both weekdays and weekend days and to include seasonal variation in the assessment.

A number of factors need to be taken into account in choosing methods for assessing physical activity:
+ Accuracy (reliability and validity)
+ Time frame
+ Nature and details of the physical activity. Physical activity is commonly considered to have three dimensions: duration, frequency and intensity (or "strenuousness").
+ The performance of an activity or group of activities such as in a job activity questionnaire.
+ Mode of data collection (personal interview, telephone interview, self-administered questionnaire or mail survey)
+ Summary estimate of physical activity for an individual (or a group), i.e., a summary estimate or score that can be used for ranking individuals according to level of physical activity)

Intensity of physical activity

Many terms have been used to characterize the intensity of physical activity such as light, moderate, hard or very hard. These terms can be related to the absolute amount of energy expenditure or oxygen consumption associated with specific types of physical activity. An example is the amount of oxygen consumed in walking at 5 km/h, which may be expressed in multiples of resting oxygen consumption. One metabolic equivalent (MET) is set at 3.5 mL of oxygen consumed per kilogram body mass per minute – an amount associated with sitting in a rested state. Walking at 5 km/h requires about 3 METs or 630 mL of oxygen consumed per minute by a 60-kg person. It should be noted that the amount of oxygen consumed by a young or old 60-kg person while walking at 5 km/h will have the same absolute level, but this constitutes a greater relative demand for the older person because maximal oxygen consumption declines with age.

Another way of characterizing intensity of activity is to describe the effect of participating in a specific type of activity relative to an individual's maximal oxygen consumption. This approach has been valuable in prescribing safe levels of exercise for cardiac patients and for otherwise healthy adults in whom exercise may produce untoward consequences (American College of sports Medicine, 1990). Because oxygen consumption and heart rate during physical exercise are highly correlated, the percentage of maximal heart rate has often been used to reflect the relative effect on maximal oxygen consumption (American College of Sports Medicine, 1990).

Finally, on a subjective level, a person may rate the intensity of an activity him/herself. For example, a person may be asked to report the frequency and duration of participation in vigorous activity that results in increased breathing or heart rate. As noted above, this would closely parallel the relative demands on a person.

Summary scores in physical activity measures

Physical activity can be described in a variety of ways, as evidenced by the many different summary scores used in various countries for large representative surveys or as part of epidemiological studies, notably in relation to monitoring cardiovascular disease risk. One way of characterizing physical activity is estimation of energy expenditure, expressed either as total energy expenditure (kilojoules, kJ) per day or week, or relative to an individual's body mass (kJ/kg/day). In addition, specific patterns of frequency and duration of physical activity can be established to clarify the particular categories and specific amounts of behaviours being reported. Some combinations of parameters have been used to reflect energy expenditure (e.g., regular moderate physical activity such as walking five times per week for 30 minutes). Another pattern reflects the

Figure 8 Examples of leisure-time physical activity (cross-country skiers and canoeists)

activity needed to improve or maintain aerobic capacity (e.g., vigorous activity performed three times per week and lasting for 20 minutes) – the traditional exercise prescription for cardiovascular disease prevention (Caspersen et al., 1994). Other summary scores have been created by combining data on the frequency, duration and intensity of all or even specific types of physical activity. Such scores have often been inferred to reflect energy expenditure.

Critical assessment of the various measures used

Measures of cardiorespiratory fitness have been considered as a surrogate of physical activity and have been employed in some prospective studies of cancer mortality (Oliveria et al., 1996). However, because genetic aspects dictate both cardiorespiratory fitness level and its response to physical activity, and because fitness measures are influenced by age, gender and other health habits, cardiorespiratory fitness level is inadequate as a measure of physical activity (Ainsworth et al., 1994).

These measures have rarely been used in epidemiological studies of cancer, and are not considered further.

Job classification

Because employed adults spend many hours at work and have the opportunity for considerable expenditure of energy, occupational titles and related approaches have been used to define levels of physical activity in cancer studies. In a study of breast cancer, Calle et al. (1998a) asked about the current and longest-held job over a woman's lifetime and identified and compared separate risks for 13 occupational groups, using housewife as a reference group. Another study used occupational titles to compare cancer risks for four occupational groupings as part of three industries, using farmworkers/agriculture as the referent group (Hsing et al., 1998a). Some studies have gone beyond using titles alone to further delineate tasks, by using resources such as the US Department of Labor's Dictionary of Occupational Titles (1993), which uses the intensity and duration of lifting,

pushing and pulling, the body's position during exertion, as well as the estimated rate of energy expenditure for work tasks, to create sedentary, light, medium and heavy groupings (Coogan et al., 1997; Coogan & Aschengrau, 1999). Others have used similar approaches to create three (Fredriksson et al., 1989) or five occupational activity groupings (Moradi et al., 1998; Bergström et al., 1999), while Dosemeci et al. (1993) created two separate indices, one for job-related energy expenditure (< 8, 8–12, >12 kJ/min) and one for sitting time (< 2, 2–6, > 6 h/day), to compare cancer risks in their population. While occupational titles and related classifications can be valuable, they are unreliable if misclassification is known to exist, or in populations where work-related energy expenditure is less prevalent. In such instances, other types of physical activity must be assessed and in some cases combined with occupational measures to yield measures of exposure.

Questionnaires

Recall questionnaires for occupation, leisure, household or other activity generally require less respondent effort and are less likely to affect the respondent's physical activity. Because recall questionnaires have time frames of one week (Sallis et al., 1985) to one year (Taylor et al., 1978) or even a lifetime (Friedenreich et al., 1998), the respondent may have to expend considerable effort in remembering details of past participation in physical activity. Recall questionnaires can be either self- or interviewer-administered, the latter entailing more interviewer training, quality control and study costs. Recall questionnaires may ask for precise details on physical activity or may solicit general reports of usual participation in physical activity over a given time frame; they can generally be characterized as global, single-item and comprehensive questionnaires.

Figure 9 Rice planting in the Philippines

Global questionnaires require comparison of one's physical activity with that of other people in general. The global self-report is very easy to use and has at least some evidence of validity (Sternfeld *et al.*, 2000). Several cancer studies have used global questionnaires (Andersson *et al.*, 1995; Kotake *et al.*, 1995; Neugut *et al.*, 1996; Marcus *et al.*, 1999; Verloop *et al.*, 2000). However, it has been questioned what precise physical activity profiles form the basis of comparison when groups differing in age, gender, or racial/ethnic status report the same self-assessment rating (Sternfeld *et al.*, 2000).

Single-item questionnaires allow rapid assessment of general patterns of physical activity. For example, among Iowa women aged 55–69 years, physical activity was measured as the weekly frequency of moderate-intensity and vigorous activity and responses to the two questions were used to group women into low-, moderate- and high-activity categories (Mink *et al.*, 1996;

Moore *et al.*, 2000a), though the repeatability and validity of these two items have yet to be reported. In prospective studies of physicians in the United States, investigators assessed the frequency of vigorous physical activity likely to promote sweating (Lee *et al.*, 1997a; Liu *et al.*, 2000). Correlations between answers to the sweat question and oxygen uptake were reported to be 0.54 for males, 0.26 for females and 0.46 for the total group (Siconolfi *et al.*, 1985).

The Godin questionnaire assesses the frequency per week of exercise lasting at least 15 minutes or more for three categories of effort: strenuous (heart beating rapidly), moderate (not exhausting) and mild (minimal effort) (Godin & Shephard, 1985). For each category, examples of activities that produce the level of effort are proposed, while an additional question asks "how often do you engage in any regular activity long enough to work up a sweat?". Coefficients for 2–4-week repeatability have ranged from 0.24–0.48 for light effort to 0.84–0.94 for strenuous effort

and 0.69–0.80 for the sweat question (Godin & Shephard, 1985; Jacobs *et al.*, 1993). Correlations of 0.54–0.61 with other physical activity surveys (Miller *et al.*, 1994), with indices of maximal cardiorespiratory fitness (0.52–0.57) and with body fat (−0.43) (Jacobs *et al.*, 1993) constitute indices of validity for this approach. Friedenreich and Rohan (1995) modified the Godin questionnaire to assess total time per week at each level of effort and multiplied each by 5-, 7.5-, and 10 kcal/min to create a summary score in kcal/week. Thune *et al.* (1997) and Gram *et al.* (1999) used one question for work and one for leisure, each graded from 1 to 4, as part of the Second and Third Tromsö Studies (Thune *et al.*, 1998). These two questions demonstrated some aspects of validity (Wilhelmsen *et al.*, 1976; Holme *et al.*, 1981; Løchen & Rasmussen, 1992). Many other single-item questionnaires have been used in cancer studies (Garfinkel & Stellman, 1988; Gerhardsson de Verdier *et al.*, 1990a; Hirose *et al.*, 1995; Fraser & Shavlik, 1997; Hartman *et al.*, 1998; John *et al.*, 1999; Terry *et al.*, 1999; Stessman *et al.*, 2000), but neither their repeatability nor validity have been reported.

Comprehensive questionnaires. In studies of college alumni, Lee *et al.* (1999a) used a recall questionnaire to assess the distance and pace of walking, flights of stairs climbed, and frequency and duration of sports or recreational activities typically performed during the past year, in order to create a kilocalorie summary score. For this questionnaire, the short-term repeatability (four weeks) for the total score was 0.76 (Washburn *et al.*, 1991), while Rauh *et al.* (1992) found two-week repeatability to be greater for flights climbed (0.68) and sports participation (0.67) than for blocks walked (0.23), again revealing better recall of participation in more intense activity. Long-term repeatability

Figure 10 Workers in a park in Hanoi, Vietnam

(9–12 months) was between 0.50 and 0.73, with coefficients ranging between 0.39 and 0.42 for flights climbed, between 0.30 and 0.54 for blocks walked and 0.63 for sports (LaPorte et al., 1983; Jacobs et al., 1993). As indices of validity, the College Alumni questionnaire has shown favourable correlations (> 0.50) with other instruments (Albanes et al., 1990) and reasonable correlations with a four-week activity history (0.31), indices of cardiorespiratory capacity (0.52) and body fat (−0.30) and motion sensor counts (0.30) (Jacobs et al., 1993).

As part of the Framingham Study, Dorgan et al. (1994) used an interviewer-administered recall questionnaire to assess hours spent per day in sleep, work and extracurricular activities in a typical day, and then multiplied the time estimates by weighting factors of 1.1, 1.5, 3.4 and 5.0, corresponding to increasing levels of oxygen uptake. Modest repeatability coefficients of 0.30–0.59 between reports two to three years apart have been found for this instrument (Garcia-Palmieri et al., 1982), while validity has been assessed by correlations with other physical activity questionnaires (0.48–0.72) and with total

energy intake (0.43) (Albanes et al., 1990).

Le Marchand et al. (1997) used the Stanford seven-day physical activity recall questionnaire (Sallis et al., 1985) to assess usual work and leisure averaged over a three-year period. The Stanford recall questionnaire asks subjects to report time spent in sleep, moderate, hard and very hard activities for each of the past five weekdays and two weekend days, differentiating between work and leisure. Repeatability coefficients for the total report have ranged from 0.34 (Jacobs et al., 1993) to 0.86 (Gross et al., 1990), while subtotals for very hard activity and sleep have correlations as high as 0.86 and 0.76, respectively (Rauh et al., 1992). However, moderate-intensity activity had correlations of 0.08–0.12 (Sallis et al., 1985; Jacobs et al., 1993) to 0.52 (Rauh et al., 1992). Validity coefficients for weekend and weekday physical activity logs were as high as 0.70 and 0.75 for moderate-intensity activity and 0.66 and 0.39 for hard and very hard activity, respectively (Taylor et al., 1984). Validity correlations were low (having an absolute value of 0.36 or less) between the survey score

and a four-week activity history, indices of cardiorespiratory capacity and body fat, and motion sensor counts (Jacobs et al., 1993).

Sandler et al. (1995) used a telephone-administered version of the Baecke questionnaire (Baecke et al., 1982). This particular physical activity recall questionnaire consists of sections for work activity, sports activity and non-sports leisure activity. Each section has questions scored on a five-point Likert scale ranging from "never" to "always" or "very often", except that for reports of the two most frequently played sports, the number of months per year and hours per week of participation are solicited. The repeatability, in terms of correlation coefficients, has been between 0.70 and 0.90 for the three sections, regardless of the study population of men and women, and whether for short (1–5 months) (Baecke et al., 1982; Jacobs et al., 1993) or longer time periods (11 months) (Pols et al., 1995). Validity for the Baecke questionnaire has been demonstrated by favourable correlations with data from other physical activity questionnaires (0.56–0.78) (Albanes et al., 1990; Miller et al., 1994) and physical activity diaries (0.33–0.66) (Pols et al., 1995; Richardson et al., 1995). Richardson et al. (1995) reported correlations of 0.46 and 0.57 comparing the survey results with cardiorespiratory fitness, and of −0.51 and −0.30 with indices of body fat for women and men, respectively.

The CARDIA (Coronary Artery Risk Development in Young Adult Study) questionnaire (Jacobs et al., 1989) consists of a set of 13 activity categories (eight for strenuous and five for non-strenuous activities). For each set, the respondent provided the total number of months of participation, with further probing for the number of months in which activities were performed for at least one hour, and additionally for activities performed for at least two hours, during any of the preceding

12 months. Slattery and Jacobs (1995) found, for a subsample of 81 CARDIA participants, that recalled activity two to three years earlier was highly associated with activity reported during the examination of the corresponding time. Spearman correlations were 0.84 for vigorous, 0.64 for moderate, and 0.81 for total activity. Distantly recalled activity was less well associated with contemporary activity reports, with correlations being 0.57, 0.45 and 0.59, respectively, suggesting that activity was not completely stable over time. As an index of validity, Slattery and Jacobs (1995) found that distantly recalled activity was significantly correlated with resting pulse rate as measured during the examination at the corresponding time (−0.21). The CARDIA questionnaire has been used in several cancer studies (Slattery et al., 1997a, 1999).

Comprehensive questionnaires have been used in large prospective studies of health professionals. Men participating in the Health Professionals Follow-up Study recalled the average time per week spent over the past year in six vigorous (jogging, running, bicycling, swimming, tennis/squash/racket-ball, calisthenics/rowing) and two non-vigorous activities (walking/hiking, flights of stairs climbed) (Giovannucci et al., 1995, 1998). Time spent in each activity was multiplied by a specific intensity code (Ainsworth et al., 1993) to derive an estimate of energy expenditure in physical activity (MET-hours/week). Two-year repeatability coefficients were 0.52 for vigorous, 0.42 for non-vigorous and 0.39 for lack of activity, while, as an index of validity, correlations with data from four one-week seasonal diaries were 0.58, 0.28 and 0.41, respectively (Chasan-Taber et al., 1996). The close correspondence between the averages and distributions for the four diaries and the recall suggested that respondents had effectively incorporated seasonal information over the preceding year to create the average weekly estimate as

part of their recall. In the Nurses' Health Study (in even years, beginning in 1986) and the Nurses' Health Study II (in 1989), women completed a physical activity questionnaire similar to that used by men in the Health Professionals Follow-up Study (Giovannucci et al., 1996; Martinez et al., 1997; Rockhill et al., 1998, 1999). The two-year repeatability of the Nurses' Health Study II questionnaire was 0.59 for activity and 0.52 for inactivity, while validity coefficients, as assessed with the four one-week seasonal diaries, were 0.56 and 0.41, respectively (Wolf et al., 1994).

The European Prospective Investigation into Cancer and Nutrition (EPIC) included physical activity recall among other lifestyle assessments in its core questionnaire (Riboli & Kaaks, 1997). Subjects were asked to describe their participation in various types of activity: work (sedentary, standing, manual, heavy manual); walking, cycling, gardening, do-it-yourself activity, exercise, and housework (hours/week); stairs climbed (flights/day). Five-month and 11-month repeatability coefficients for energy expenditure were 0.86 and 0.75 for men and 0.63 and 0.68 for women, respectively (Pols et al., 1997). The correlation between energy expenditure estimated from the questionnaire with that from four three-day seasonal diaries was 0.43 for men and 0.51 for women.

Histories

Quantitative history. The retrospective quantitative history is the most thorough physical activity enquiry because it requires detailed recall of physical activity for time frames of up to one year or longer. The immense detail associated with the quantitative history implies a very large memory burden for the respondent and has high costs for administration, interviewer training, quality control and data-processing (Caspersen, 1989). Quantitative history questionnaires that cover at least one

year can assess seasonal physical activity. For example, McTiernan et al. (1996) asked respondents to report months, average frequency, and duration of participation for individual activities, identified from listed activity groupings, over the preceding two years, modelling their questionnaire on the Minnesota Leisure-Time Physical Activity Questionnaire (MLTPAQ) (Taylor et al., 1978). The time reported for each activity in the MLTPAQ was multiplied by an activity metabolic index (AMI, work metabolic rate divided by basal metabolic rate) to create a summary score approximating energy expenditure per week. The summary score has been further reported according to total, light (AMI < 4), moderate (AMI 4–6), and heavy (AMI > 6) activity following a convention that applies almost solely to men aged 35–57 years participating in that study. The short-term repeatability (about four weeks) for total, light, moderate and heavy activity was 0.92, 0.73, 0.80 and 0.95, respectively, while values for long-term (one year) repeatability were 0.69, 0.60, 0.32, and 071, respectively (Richardson et al., 1994). As a reflection of validity, the MLTPAQ has compared favourably with other physical activity questionnaires, showing correlations of greater than 0.47 (Albanes et al., 1990). In addition, the MLTPAQ has shown reasonable correlations with a physical activity record (0.47), an index of aerobic capacity (0.47) and body fat measurements (−0.24) (Richardson et al., 1994).

Lifetime history. To overcome part of the problem of memory burden, one lifetime physical activity questionnaire (Friedenreich et al., 1998) used cognitive interviewing techniques for women aged 34–65 years, noting a high test–retest repeatability (Pearson correlation ≥ 0.72) for a 6–8-week repeat questionnaire for occupational, household, exercise/sports and total activities for women. Repeatability was generally highest for
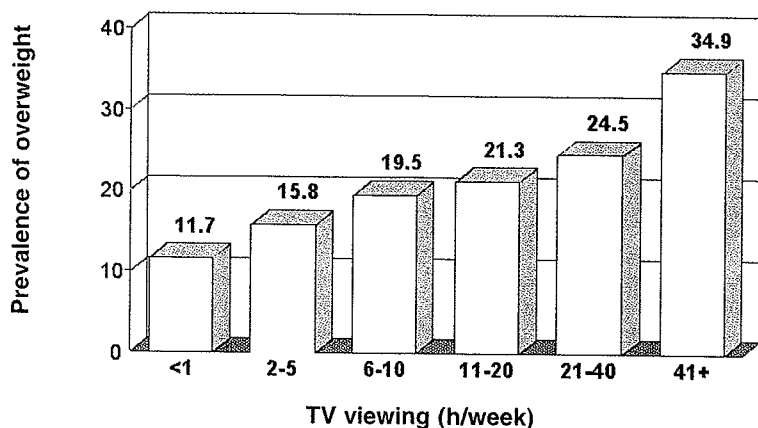
Figure 11 Relationship between overweight and TV viewing
Data from Ching et al. (1996)

occupational activity, for high-intensity activities, and for each type of activity reported for time periods earlier than for the year preceding the interview. On the other hand, repeatability of lifetime physical activity recall may vary by gender. A comparison of recalled physical activity with data from essentially the same questionnaire administered 30–35 years earlier found intra-class correlations of 0.43 for light and 0.45 for moderate weekday (occupational) activity for both men and women, and 0.38 for hard free-day (leisure) activity for women only (men had a correlation of 0.26) (Falkner et al., 1999). In addition, while under-estimation occurred for both men and women who reported time spent in past weekday activities; only men over-estimated hard free-day activity. Otherwise, estimated time did not vary by gender when examined by intensity level of activity. Cognitive interviewing techniques hold promise for lifetime quantitative questionnaires in terms of both the quantity and precision of responses, compared with traditional standardized interviews, but they entail additional costs for interviewer training, time spent in conducting interviews and difficulties in coding the responses. While they may be an excellent way to assess activity in a specific population, their specificity may limit the possibility of comparing different populations or use in other settings (Fisher et al., 2000).

Concluding comments

Physical activity takes many forms, which explains, in part, why it is difficult to measure and why it has been measured in so many ways as part of experimental, intervention and epidemiology research. While some measures such as calorimetry and use of doubly labelled water are precise, they tend to restrict the physical activity behaviour being measured. On the other hand, job classification and recall measures (global, general, comprehensive and quantitative histories) are less precise but do not restrict physical activity measurement and are most often used for epidemiology and survey research.

One area of particular interest is sedentary behaviour as a distinct entity.

Sedentary behaviour needs to be seen as different from simple non-participation in physical activity. There is evidence that time spent in sedentary behaviour (particularly television viewing) is significantly associated with being overweight, even among people who are quite physically active in their leisure time (Figure 11) (Salmon et al., 2000), but data on the health impact of such behaviour, its prevalence in populations and trends over time are sparse (Jebb & Moore, 1999; Pratt et al., 1999; Owen et al., 2000).

Many questionnaires that have been used in cancer epidemiology have been tested for reliability and validity. In general, strenuous physical activity appears to be recalled with greater accuracy than either light or moderate-intensity activity, while recent activities are recalled more accurately than those performed at earlier times. However, so many different samples of persons varying in age, gender and other important sociodemographic characteristics have been studied, so many different validation measures have been applied, and so many different correlational statistics have been used to assess their quality that it is not possible to determine what type of questionnaire is best or even qualitatively better than another.

A programme of new research to establish a broader set of reliable and valid measures is gaining momentum (Macera & Pratt, 2000). The use of objective tools such as accelerometers (motion sensors), heart rate monitors and direct observation of physical activities is also being explored (Sallis & Saelens, 2000). These assessment procedures may help to validate the self-report measures that are used because of the practical and financial constraints inherent in carrying out large-scale population surveys.