

From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

PHYSICAL ACTIVITY FROM THE EPIDEMIOLOGICAL PERSPECTIVE – MEASUREMENT ISSUES AND HEALTH EFFECTS

Ylva Trolle Lagerros



**Karolinska
Institutet**

Stockholm 2006

All previously published papers were reproduced with permission from the publishers.

Published and printed by Repro Print AB, Stockholm, Sweden

Cover photo by Jacob Forsell. The National March 1997.

© Ylva Trolle Lagerros, 2006

ISBN 91-7140-681-6

Till minnet av farmor

SUMMARY

The aim of this thesis was to increase our understanding of links between physical activity and health. The relationships between physical activity and breast cancer risk and between current physical activity and overall mortality were examined. Additionally, a novel method was developed to measure physical activity in epidemiological studies. Besides validation studies, a large cohort was enrolled for assessment of physical activity with this method.

Although an association between physical activity during adolescence and breast cancer is biologically plausible, results from individual studies have been inconclusive. This led us to conduct a quantitative summary analysis of published studies. Nineteen case-control and four cohort studies were included in our meta-analysis. Comparing the highest to the lowest category of physical activity, the summary relative risk (RR) of breast cancer was 0.81 (95 % confidence interval [CI] 0.73–0.89). This risk reduction of almost 20 % was fairly consistent across different strata. The analysis identified an important source of heterogeneity within the literature; the various methods used to assess physical activity, a complex exposure.

In a population-based cohort study of 99,099 Norwegian and Swedish women, we found that current rather than past physical activity substantially reduces mortality. This was observed even at low levels of physical activity, and was accentuated with increased physical activity. During an average 11.4 years of follow-up, risk of death decreased monotonically over five categories of physical activity at the time of enrollment (p for trend <0.0001) and was reduced by half in the highest compared with the lowest category (RR = 0.46; 95 % CI 0.33–0.65). Physical activity was assessed on a 5-point scale, which makes it difficult to translate the results to a quantifiable public health message.

A prerequisite for advancements in our understanding of health effects of physical activity is better assessment methods. We developed an instrument of self-reported time spent on different intensity levels of physical activity (and inactivity) during a typical day, allowing for estimation of total energy expenditure on an interval scale. In a first validation study, we tested if 80 volunteers using our instrument could correctly estimate MET (Metabolic Energy Turnover) values during concurrent work on a bicycle ergometer. The Pearson correlation coefficient between true and estimated METs was 0.89 (range 0.81 – 0.95). In a retest reliability assessment of 20 subjects, intraclass correlation coefficient was 0.99.

In a second validation study we addressed the ability to correctly remember and integrate “usual” energy expenditure over time. A population-based sample of 418 Swedish men and women, aged 20–59 years, completed a questionnaire containing the new instrument. For validation, three 24-hour recalls by phone served as the gold standard. Reproducibility was assessed through administering the instrument another three times over 6 months. Pearson correlation between usual daily energy expenditure measured by the instrument and the 24-hour mean recall was 0.73.

Reproducibility showed an intraclass correlation of 0.55. We concluded that the instrument provides reasonably valid estimates of total energy expenditure.

We established a general population cohort based on 43,876 subjects who took part in The National March, a fund-raising event for the Swedish Cancer Society. All participants were invited to fill out a 36-page questionnaire, of which three pages were devoted to physical activity. We cross-sectionally compared measures of physical activity obtained through different means of inquiry. Household and leisure time physical activity represented no more than 17 % of total physical activity assessed by the new instrument. The Spearman correlation coefficient between total physical activity obtained with the new instrument and the sum of household and leisure time physical activity was 0.26. The correlation was even lower when comparing total activity obtained with the instrument to self-rated fitness and self-rated total activity judged relative to peers, indicating that the estimated physical activity level in an epidemiological study is contingent on the mode of inquiry.

Keywords: Breast neoplasms, cohort studies, energy metabolism, epidemiologic methods, exercise, meta-analysis, mortality, questionnaires, review, validation studies, women

1. LIST OF ORIGINAL PAPERS

This thesis is based on the following papers which will be referred to by their Roman numerals (I-V).

- I. **Lagerros YT**, Hsieh SF, Hsieh CC.
Physical activity in adolescence and young adulthood and breast cancer risk: a quantitative review
Eur J Cancer Prev. 2004 Feb;13(1):5-12.
- II. **Trolle-Lagerros Y**, Mucci LA, Kumle M, Braaten T, Weiderpass E, Hsieh CC, Sandin S, Laggiou P, Trichopoulos D, Lund E, Adami HO.
Physical activity as a determinant of mortality in women
Epidemiology. 2005 Nov;16(6):780-5.
- III. **Lagerros YT**, Bergström R, Nyrén O.
Validity of perceived work intensity using a novel instrument
Submitted
- IV. **Lagerros YT**, Mucci LA, Bellocco R, Nyren O, Balter O, Balter KA.
Validity and reliability of self-reported total energy expenditure using a novel instrument
Eur J Epidemiol. 2006;21(3):227-36.
- V. **Lagerros YT**, Bellocco R, Adami HO, Nyren O.
Assessments of physical activity in epidemiologic studies are sensitive to the method of inquiry
Submitted

Previously published papers were reprinted with kind permissions of Lippincott Williams & Wilkins (I, II) and Springer Science and Business Media (IV).

CONTENTS

1.	LIST OF ORIGINAL PAPERS.....	vi
2.	LIST OF ABBREVIATIONS.....	ix
3.	INTRODUCTION.....	2
4.	BACKGROUND.....	3
	Conceptual Framework.....	3
	Intensity.....	4
	Duration.....	5
	Frequency.....	6
	Methods to assess PA and EE.....	6
	Objective methods.....	6
	Self-report methods.....	9
	Physical Activity and Cancer.....	12
	Biological plausibility.....	12
	Breast cancer.....	13
	Other site-specific cancers.....	14
	Physical Activity and Mortality.....	14
5.	AIMS.....	15
6.	SUBJECTS AND METHODS.....	16
	Study designs, Populations and Physical Activity Assessments.....	17
	Study I – Physical activity in a meta-analysis.....	17
	Study II – Measuring physical activity the rudimentary way.....	19
	Study III – Development of an instrument, a first validation study.....	20
	Study IV – A second validation study in a random sample.....	23
	Study V – Using the novel instrument in a cohort study.....	24
	Statistical methods.....	26
	Study I.....	26
	Study II.....	28
	Study III.....	30
	Study IV.....	31
	Study V.....	32
7.	RESULTS AND COMMENTS.....	33
	Study I.....	33
	Study II.....	34
	Study III.....	36
	Study IV.....	37
	Study V.....	38

8. GENERAL DISCUSSION	39
Methodological Considerations	39
Validity	39
Precision	44
Interpretations and Implications	45
9. ONGOING AND FUTURE RESEARCH.....	48
10. CONCLUSIONS.....	50
11. SAMMANFATTNING (SUMMARY IN SWEDISH).....	51
12. ACKNOWLEDGEMENTS.....	54
13. REFERENCES.....	57

2. LIST OF ABBREVIATIONS

BMI	Body Mass Index
CI	Confidence Interval
EE	Energy Expenditure
g	gram
h	hour
HR	Hazard Ratio
kcal	kilocalorie
I	Intensity
IARC	International Agency for Research on Cancer
J	joule
MeSH	Medical Subject Headings
MET	Metabolic Energy Turnover
SE	Standard Error
PA	Physical Activity
PAF	Population Attributable Fraction
RMR	Resting Metabolic Rate
RR	Relative Risk
t	Time, duration
TEE	Total Energy Expenditure
TEF	Thermic Effect of Feeding
URL	Uniform Resource Locator
VAS	Visual Analogue Scales

3. INTRODUCTION

Industrial progress and increased employment in the service sector generally have resulted in less occupational physical activity, while at the same time modern technology has also made it increasingly convenient to remain sedentary. Many people lead a life with little or no physical activity and their leisure time is often spent on sedentary activities such as surfing the internet, playing computer games and watching television. Sixty per cent of the world's population is estimated to lead a sedentary life.

¹ Simultaneously, society is facing new patterns of illness in the multidimensional room of age, physical activity and body measures. While it is clear that physical activity has a far reaching influence on health, many questions remain to be answered – questions that stem from crucial health concerns that need to be addressed.

Epidemiology is the study of the distribution and determinants of health-related states in specified populations and the application of this study to monitor health problems. ² It is often considered the core science of public health. Although epidemiology started with the study of contagious infectious diseases during the 19th century, it was during the second half of the last century that epidemiological concepts and methods became more widespread and a broader range of health problems were studied.

Contrary to infectious disease epidemiology, where time between infection and disease is short which facilitates the search for causal components, epidemiological studies on physical activity face the challenge of exposures possibly from decades ago. These past exposures may play a role in the development of disease, or perhaps only in conjunction with other causal components, such as hormones.

One of the first epidemiologic studies on the impact of physical activity was done in 1962. In a study of railroad employees, those who were sedentary were found to have higher death rates than physically active employees. ³ The results could have been confounded by unmeasured factors; for instance, the well paid clerks were more likely to smoke than other personnel. It was not until two years later that the Surgeon General at the U.S. Department of Health, Education and Welfare, declared cigarette smoking and lung cancer causally related.

Since the sixties, there has been an explosion of studies focusing on the topic of physical activity and health. Despite the fact that physical activity has gained increasing attention, the lack of practical, valid, reliable and sensitive instruments for self-recording of all physical activity and inactivity has been a limiting factor in this important area of research. With better methods for exposure quantification, we can reveal not only the causal link between exposure and disease, but also how the exposure might act as a confounder or an effect measure modifier of various risk factor/disease relationships.

The aims of this thesis were to investigate the association between physical activity and later illness (breast cancer risk and all cause mortality among women) and to develop and validate a novel instrument for the self-reporting of total physical activity suitable for epidemiological studies.

4. BACKGROUND

CONCEPTUAL FRAMEWORK

Physical activity is a multidimensional and complex exposure to measure. Research on physical activity is done in a wide range of disciplines and from different perspectives, which sometimes results in conceptual confusion.

Physical activity (PA) is defined as “any bodily movement produced by skeletal muscles that results in energy expenditure.”⁴ Daily total energy expenditure (TEE) is a result of posture, spontaneous and voluntary physical activity (EEPA), resting metabolic rate (RMR), and the thermic effect of feeding (TEF)^{5,6} (which is the energy expenditure needed for digestion, absorption, and the increase of sympathetic nervous system activity after eating a meal). See figure 1. TEE of an activity is sometimes called the gross cost of the activity, while the net cost, is the cost of the activity by itself,⁷ subtracting RMR and TEF. A stable body weight is maintained when TEE balances total energy intake.

In physics the standard unit of energy is joule (J), but in the world of energy metabolism, the unit most commonly used is the calorie (or kilocalorie, kcal, which is equivalent to 1,000 calories). One calorie corresponds to approximately 4.19 J.

RMR have been extensively studied and is constant within and between persons. Besides age-related changes it only varies 5–10 per cent in adult life, and comparisons made within age, sex and weight groups show that 85 per cent have a RMR within 10 per cent of the mean.⁸ Physical activity is the factor that can increase TEE the most.

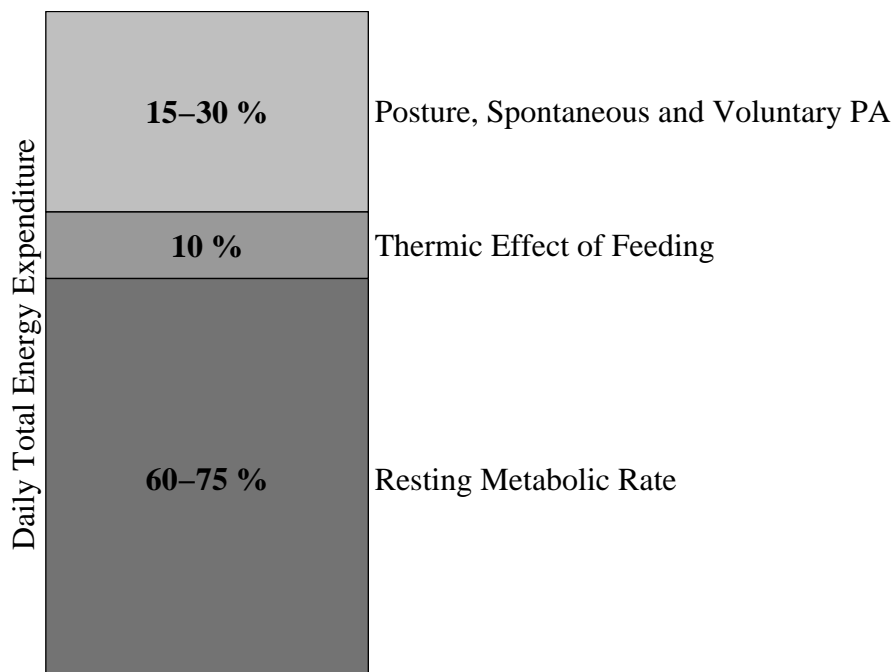


Figure 1. Components of daily total energy expenditure. $TEE = EEPA + TEF + RMR$. The energy expended on physical activity is the most variable component.

Physical activity and energy expenditure are not synonymous, but many researchers extrapolate measures of physical activity to units of energy expenditure before analyzing their studies.⁹

Optimally, epidemiological studies should identify all body movements and obtain information on dose – intensity, duration and frequency – and maybe even purpose of movement. With information on all these dimensions, comparison of results across studies would be more informative. Unrevealed dose-response relationships could be discovered and epidemiologists would be able to supply public health professionals with evidence useful in their missions. Results from studies inquiring solely about one dimension can not easily be converted to public health recommendations.

Intensity

The concept of intensity (strenuousness or power) has been defined in different ways in the interdisciplinary field of physical activity research. Here intensity (I) is defined as the power consumption (energy consumption per unit of time) or $I = EE / t$, where t is the duration of the activity. Absolute work intensity is measured in Watts, but it is often more convenient to consider intensity per body mass (W/kg) or metabolic energy turnover (MET) which is multiples of RMR. For the average adult, 1 MET (1.16 W/kg) corresponds to an energy expenditure of 1 kcal per kg body mass per hour or the approximate oxygen consumption of 3.5 ml O₂ per kg body mass per minute. See figure 2.

Intensity in questionnaires

Obtaining absolute intensity from questionnaires is done by assigning each activity a specific MET value, obtained from reference lists.^{10,11}

For an individual with a weight of 60 kg, snow shoveling by hand is estimated to correspond to 6 METs, or requires 360 kcal/hour or 1,260 ml oxygen/minute. These requirements are absolute, thus, they are the same for any 60 kg person, regardless of other factors such as age. However, as maximal oxygen consumption declines with age,⁷ the relative demand on the older person will increase.

When physical activity questionnaires give the respondent a few choices in terms of intensity (such as no, low, moderate, high and vigorous) one overlooks the potential problem that the perception of intensity is highly dependent on duration, age, gender and fitness. Only in a homogenous sample might the relative and absolute intensity be similar.¹²

Some questionnaires ask the respondent to report frequency and duration of activities where physiological parameters such as induced sweating,¹³⁻¹⁷ increased heart rate and/or breathlessness^{13,18-20} mark the intensity. However, the physiological response for a specific intensity is still likely to be greater for an unfit or older individual with lower cardiorespiratory fitness and less muscle mass.²¹

Expressing intensity relative to peak ability is an alternative method for characterizing physical activities. For example, The American College of Sports Medicine

recommends intensity equal to 60–90 % of one’s maximum heart rate or 50–85 % of one’s maximum oxygen uptake for 20–60 minutes, 3–5 days a week, to develop and maintain cardiorespiratory fitness, body composition and muscle strength.²² Intensity relative to peak ability is common in resistance training as well, where the key is repetition of maximal contraction force for a given muscle group.⁷ Although the use of relative intensity is well supported by experimental data, the use of absolute intensity, free from each individual’s subjective view of effort, is favored in epidemiological studies.

Duration

The response alternatives for duration are usually given as interval options with minutes or hours per day or per week. Duration is a challenge to measure. We engage constantly in physical activity or inactivity, from sleeping or working for hours, to short bursts of muscle contraction lifting something or tapping our fingers.

Intermittent activity, undertaken in short sessions, has been shown to improve cardiorespiratory fitness to the same degree as an activity of the same intensity undertaken in a longer session for the same total amount of time as the intermittent activity.^{1,23,24} It is reasonable to believe that there is a similar association between intermittent activity and health as well. Short bouts of activity, such as walking up the stairs instead of taking the elevator and playing with children are important; small changes that increase daily physical activity may lead to substantial health benefits. Thus, the optimal method to measure physical activity should be sensitive to all achievements – even small ones with short durations.

A 60 kg person is watching television for one hour. What is the estimated energy expenditure?

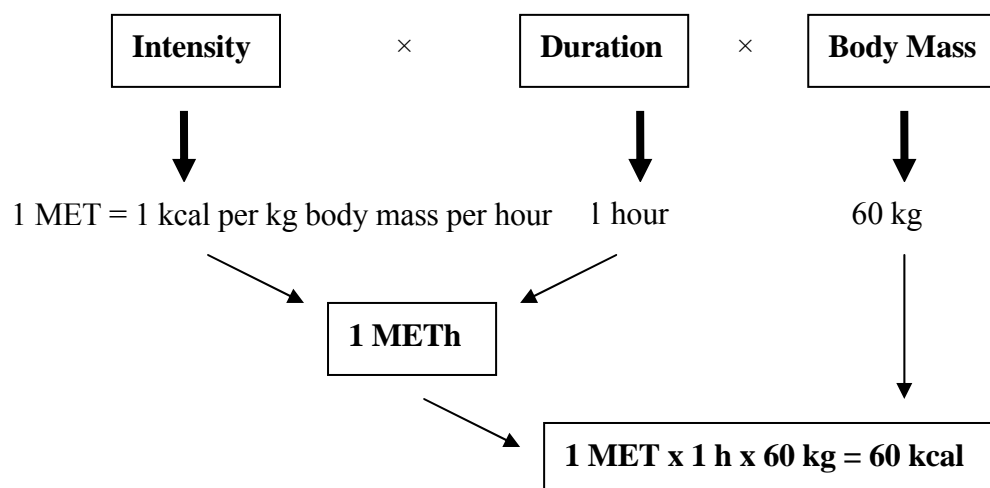


Figure 2. The volume of activity is the product of absolute intensity, duration and frequency. Volume is described in kcal, METhours or METminutes.

Frequency

With what regularity is a certain activity performed? This can be expressed as number of times a day, a week or a month. In countries where seasons vary greatly and therefore the possibility to participate in various outdoor activities, weather can become a barrier for physical activity. A number of studies have shown a relationship between seasonality and frequency of physical activities.²⁵⁻²⁹ As weekly leisure time energy expenditure can be higher in spring and summer, seasonality may be considered when planning a study or an intervention.

METHODS TO ASSESS PA AND EE

Implicit in epidemiological research is the ability to estimate the strength of an association between exposure and disease with minimal error. Accurate measurement of physical activity or inactivity is fundamental for the:

- identification of causal associations between physical activity and health outcomes
- identification and quantification of the dose-response relationships between physical activity and health outcomes
- documentation of changes in physical activity within and between individuals over time
- formulation of public health recommendations
- validation of intervention programs
- comparison of physical activity levels between populations (particularly when cultural and language differences exist between these populations)
- measurement of physical activity in children and other groups of individuals who have a limited capacity for accurate self-appraisal

As physical activity takes many forms, it has been measured in a variety of ways in experimental, interventional and epidemiological research. Methods to assess physical activity and energy expenditure could be divided into subgroups, as seen in table 1. Choosing the method is a balance between time, cost and validity. All methods have limitations, but all are useful in particular instances.

Objective methods

The objective methods based on biological and physiological approaches (e.g. heart rate monitoring, accelerometry and doubly labeled water) are harder to apply in large population studies than self-report assessments. Typically these methods have been restricted to relatively small sample sizes (i.e., <1,000 individuals). Some of these methods, such as heart rate monitoring and accelerometry are currently used in larger studies. Objective methods have also been used to validate other subjective physical activity assessment methods.

Table 1. Some methods used to measure physical activity and energy expenditure.

Method	Objective	Self-reported	Physical activity	Energy Expenditure
Doubly Labeled Water	×			×
Calorimetry	×			×
Heart Rate Monitoring	×			×
Ventilation	×			×
Cardiorespiratory Fitness	×		×	
Body Temperature	×			×
Motion sensors	×		×	
Behavioral Observation	×		×	
Psychophysical Rating Scales		×	×	
Records		×	×	
Logs		×	×	
Recalls		×	×	
Questionnaires		×	×	

Physiological approaches to measure energy expenditure

Since more than 95 % of the energy expended by the body is derived from the reaction of oxygen with nutrients,⁸ an individual's metabolic rate can be calculated once VO_2 is known. Below are some different approaches to assess VO_2 to estimate energy expenditure or physical activity.

- Doubly labeled water (DLW) is a form of indirect calorimetry and is frequently considered the gold standard to estimate total EE. As indicated by the name, DLW is water-based, consisting of the stable isotopes 2H_2O and $H_2^{18}O$, and it is consumed by the study subject.⁹ The isotopes distribute themselves evenly throughout the body,³⁰ and are gradually secreted in the subject's urine. Depending on the isotope dose and excretion rate, the latter of which is dictated by environmental temperature and the subject's activity level, the urine collection period usually spans between one and two weeks. The rates at which the isotopes are eliminated from the body are proportionate to the degree of metabolic CO_2 (VCO_2) production. Thus, oxygen uptake (VO_2) and TEE can be calculated for the study period from the difference in the elimination rates of the isotopes.^{6,9,30} This method is safe, precise, and non-invasive, and can, for example, be used in children^{31,32} and pregnant women.³³ DLW is particularly useful for the assessment of TEE in free-living conditions, as no monitors are worn, and is thus particularly appealing for use in children. However, the isotopes and the measurement methods are expensive, considering that for each dose; only one measure of energy expenditure is made. Furthermore, collection of complete urine samples is essential at the appropriate times following dosing for the method to succeed. Therefore, DLW is rarely used in large studies. Furthermore, although PAEE can be estimated using the DLW method by subtracting RMR and TEF from TEE, it cannot be used to differentiate between intensity, duration and frequency of specific activities.

- Indirect calorimetry – the participant wears a mask and carries the equipment needed for analyzing the expired air to measure VO_2 .¹ Besides the affect on behavior that wearing the equipment is likely to have (Hawthorne effect), it is cumbersome and expensive and thus not appropriate for use in epidemiology.
- Heart rate monitors – although a strong linear relationship between heart rate and VO_2 exists at higher levels of energy expenditure,⁵ this method is less precise for assessing EE at low intensities. Furthermore, other factors such as emotional stress, body temperature and medication also influence heart rate.⁹ Despite this, heart rate monitoring has worked well in medium sized epidemiological studies.^{34,35}
- The close relationship between ventilation and VO_2 has led to the development of devices to measure ventilatory response to physical activity,³⁶ but these methods are yet to prove applicable in large scale studies.
- Cardiorespiratory fitness (the ability of the cardiovascular and respiratory systems to supply oxygen to the working muscles)⁷ is determined by exercise tests and correlates highly with maximal VO_2 . Fitness may be less prone to misclassification than other self-rated measures.³⁷ and is sometimes used as a measure of physical activity in epidemiological studies.³⁸ Although fitness and total physical activity are correlated,³⁹ they also have independent components.^{40,41} Fitness is complex as it is influenced by age, gender and other habits and genetics play an important role in how well physical fitness responds to training.^{30,42}

Since almost all energy released by metabolism is converted to heat⁸ this can be used to calculate energy expenditure. Direct calorimetry is based on this principle. Body temperature can also be used to calculate the energy expenditure of activity, but it is inconvenient – steady-state takes time⁹ which makes it unfeasible for all but experimental studies.

Motion sensors

The word pedometer is Greek and means “foot measurement” – as it measures the distance travelled by foot. The pedometer, usually clipped to a belt or worn around the ankle, counts steps in response to the force generated by the body’s mass connecting with the ground via the foot (personal communication Paul Franks). It measures walking-related activity, but the length of a step varies with setting and different brands seem to detect steps differently,⁴³ furthermore, unfortunately ordinary life is often more than walking on a flat surface.

Accelerometers, on the other hand, can measure movement (acceleration and deceleration) in one (vertical), two (vertical, and medio-lateral), or three (vertical, medio-lateral and antero-posterior) planes.⁴⁴⁻⁴⁶ Intensity, duration and frequency can be assessed. However, some activities do not involve variations in acceleration. Isometric muscle contraction or muscular work against some external force, such as weight lifting, carrying and pushing, or activities like uphill walking, walking on soft surfaces, and swimming, bicycling, skating or rowing, are not detected well via accelerometry.⁴⁷ Thus, physical activity is likely to be underestimated using accelerometry, if activities of the nature described above are common.

Recently other portable devices to measure physical activity have been developed. Among them are a combined heart rate recorder and movement sensor⁴⁸ and a device for analyzing body motion and posture changes resulting in a detailed record of performed activity.⁴⁹ This is a promising area of research. The development of lightweight monitors with small computers that can store large amounts of data will probably make these methods to estimate energy expenditure more available to epidemiologists in the future. So far, for larger epidemiological studies, the cost of the monitors (between several hundred and several thousand USD), the Hawthorne effect, and the problem with compliance may be reasons why these methods are not yet in common use.

Behavioral observation

The labour-intensive method of watching and recording a person's activities is quite straightforward, but not the method of choice in larger studies. However, studies that base their physical activity estimate on occupation, classified by someone other than the respondent, resemble the method of behavioral observation to some extent.

Self-report methods

In 1997 the American College of Sports Medicine's journal devoted an entire supplement to more than 30 different instruments for self-reported physical activity.⁵⁰ With the growing interest in physical activity, new instruments continuously appear – most likely due to the fact that physical activity is a complex exposure and no instrument is adequate for every situation and every population.

Psychophysical rating scales

The subjective perception of exertion has thoroughly been studied by Borg *et al.* He has developed internationally popular scales for the evaluation and monitoring of exercise intensity. The RPE scale is a scale of ratings (R) for perceived (P) exertion (E). The scale steps vary from 6–20 and is linearly associated with exercise intensity and heart rate (from 60–200) during work on a bicycle ergometer.⁵¹ The category ratio scale (CR-10) is anchored at the top by the category “maximal exertion”. Thus, two individuals working at their maximal working capacities will experience the same degree of exertion although their physical outputs may be different.⁵² Based on this, other categories represent equivalent locations with respect to maximum sensation. The scales measure the subject's perceived “effort sense”, which is a type of intensity, but one that is relative to the subject's fitness level. Even if relative intensity seldom is the focus of epidemiological studies, the CR-10 scale has been used as a complement to physical activity survey questions for estimating the degree of effort when exercising.⁵³

Nonetheless, these scales are primarily used to measure subjective physical strain/fatigue when self-rating concurrent work load. Since the steps are not anchored, the scales assess exertion on an ordinal scale. The scales based on numbers and verbal expressions, lack distinct levels of absolute intensity. This makes them less useful in epidemiological studies, where estimations of energy expenditure often are preferred.

Physical activity records

Physical activity records are based on the diary idea – the study participant is asked to keep a record of the different types of activities undertaken, and the time spent doing each of them during a specific time period.^{9,30} The record is then processed using coding schemes which classify each activity by, for example, rate of energy expenditure.¹⁰ This method can detail all activities undertaken, but it is cumbersome and it takes time for the study subject and the researcher to keep the diary and decode the entries, respectively. The recording process may in itself produce changes in physical activity patterns during the time of recording. Thus, it is not the method of choice for the large epidemiological study, but it is a useful method for validation studies.

Physical activity logs

As with the physical activity record, the study participant is asked to report the time spent doing different types of activity during a given time period. Typically, physical activity logs provide a list of specific activities to choose from.⁹ The list facilitates the journal keeping for the study participant and the data is easier to process. There is a risk of losing important information as such a list can never be complete. In particular, low intensity activities, such as routine light activity, household chores and spontaneous activity tend to be underrepresented in physical activity logs. By missing the lower end of the continuum of physical activity, the instrument could suffer from floor effects. The sedentary population would be misclassified when the lowest score available is too high.⁴⁴ This method is better suited to answer a specific question by the researcher, such as the participation rate in an exercise training programme.¹ This method may also influence the participant's physical activity pattern – just like the diary.

Recalls

The recall method, contrary to records and logs, runs a lower risk of affecting the patterns being measured. The study participant is asked to recall past activity, usually in an interview, in person or by phone.⁹ The time frame could be anywhere between 24 hours, a week, a year or a life time. Skilled interviewers can obtain a good estimate of recalled activity by cueing, i.e., using questions that enhance memory capacity, and by taking a retrospective look back to allow the participant to search his or her memory for activities s/he may have forgotten to mention.^{50,54} The disadvantage of the recall method is the time and the cost of educating the interviewers, calling the study participants and coding the data.

Questionnaires

Questionnaires are, compared to other instruments, easy to administer, generally low profile, non-reactive, easily distributed and do not require a lot of motivation or time from the study participant. With a decreased investment of time and money compared to many other methods, questionnaires provide information on physical activity and other factors of importance from a greater number of study subjects. Hence, this is the method of choice in large epidemiological studies. There are many different physical activity instruments developed for questionnaires – all with different strengths and

weaknesses. By and large questionnaires can be characterized as global, single-item or comprehensive questionnaires.⁶

Global close-ended multiple choice questionnaires ask the respondent to rate their relative level of physical activity or fitness compared to others of the same age and gender. This self-report is simple and short and is used in a variety of studies, often in combination with other questions.^{14,16,55-58} Validity has been assessed against measures of leisure time physical activity or fitness.^{14,59-61} This type of question gives a measure on a scale relative to peers. Conceivably the same self-rated answer stands for different levels of activity depending on culture, or even in which social context friends are made – on the soccer field or in the chat rooms of internet.

Single-item questionnaires lack the ability to capture all activities during daily life, but give a quick estimate of some component of physical activity. Participants could, for example, rank their overall level of physical activity on a 5-point scale, as is done in one of the studies of this thesis. They could rate time spent sitting during leisure time or the time spent sitting, standing/walking, etc, on a working day.^{62,63} The frequency of activities requiring light or vigorous effort, or the question “For how many hours per week, on average, do you engage in activity strenuous enough to build up a sweat?” are examples of single-item questionnaires.^{15,17}

Comprehensive questionnaires request more in depth information than global or single-item questionnaires. Some give an extensive list of activities and ask participants to indicate the duration and frequency of the activities in which they participate, thereby enabling the calculation of energy output. These questionnaires are often modeled after the Minnesota Leisure Time Physical Activity Questionnaire. The questionnaire consists of 63 sports, recreational, yard and household activities and was originally created for an interview.⁶⁴

In the Nurses’ Health Study physical activity has been assessed in a number of ways. In a self-administered questionnaire were questions about the frequency and duration of specific sports and recreational activities, while other questions were concerned with aspects of daily life, such as usual walking speed and the number of flights of stairs climbed each week.¹⁷

The interviewer-administered CARDIA (Coronary Artery Risk Development in Young Adult Study) questionnaire⁶⁵ asks participants to specify any activity from a list of 13 activity categories that they participated in for at least one hour during the last 12 months. Leisure, work and household activities are covered and units based on METs are calculated.

Another commonly used questionnaire is the Baecke questionnaire.⁶⁶ It is self-administered and consists of three sections; work, sports and non-sports leisure time activity. Each section has several questions scored on a five point Likert scale – from never to always or very often. For the two most frequently reported sports, there are additional questions on frequency and duration.

These are examples of some commonly used questionnaires. Errors in the estimation of physical activity in epidemiological studies are indisputably substantial. Physical activity is routinely allocated less space, questions and thought than most other variables measured in epidemiological questionnaires – leading to less precision and a more uncertain outcome.

PHYSICAL ACTIVITY AND CANCER

In 1700, when the Italian physician Bernardino Ramazzini wrote his famous thesis on occupational medicine (*De morbis artificum*), he kept coming back to the fact that physical activity was central in the etiology of many diseases.⁶⁷ Although this was more than 300 years ago, the vast majority of studies within the areas of physical activity and cancer have been conducted during the last decades. A Medline search for the MeSH (Medical Subject Headings) terms “exercise” and “cancer” yields more than 3,000 articles.

Biological plausibility

Numerous mechanisms have been proposed to explain the association between physical activity and cancer. Our understanding is incomplete; carcinogenesis is a complex and long-lasting phenomenon, possibly influenced by numerous other factors such as age, gender, genetic susceptibility, fat distribution, type of cancer and stage in the cancer’s development. Since physical activity affects the body in many different ways, multiple pathways are plausible.

The most endorsed hypothesis is perhaps that of the association of endogenous reproductive hormones with physical activity. We still do not have the complete picture concerning risk factors for breast cancer, but early age of menarche, late menopause, nulliparity, lack of lactation and hormone replacement therapy increase breast cancer risk.⁶⁸⁻⁷⁰ One could draw the conclusion that cumulative exposure to endogenous sex hormones is a determinant of breast cancer risk. Several of these factors are affected by physical activity; therefore, if physical activity modifies breast cancer risk, it is likely to do so through a hormonally mediated pathway.

But hormones are not just a female concern. Prostate cancer is rarely seen in men with Klinefelter syndrome who typically have substantially decreased androgen levels.⁷¹ Likewise, athletes have lower basal levels of testosterone and it has been hypothesized that physically active men may be at a decreased risk for prostate cancer,⁷² however, physical inactivity is not currently considered a definite risk factor.⁷³

The strong association between physical activity and colon cancer gave rise to the idea of a mechanical mechanism where physical activity would accelerate colon transit time,⁷⁴ which in turn was hypothesized to decrease the contact between colonic mucosa and potential fecal carcinogens or promoters.⁷⁵ However, a number of trials have failed to show an effect of exercise on transit time.⁷⁶⁻⁷⁹

Other studies have investigated if the protective effect of physical activity could be due to improved immune function. Dhabhar *et al.*, have in a number of studies in mice models, documented that short term stress can exert immunoenhancing effects.^{80,81}

However, there seems to be a J-shaped curve, where excessive amounts of high-intensity exercise, such as running a marathon, can cause diminished immunosurveillance and increased risk of infection.⁸²⁻⁸⁴

Several studies have shown that high circulating levels of insulin-like growth factors (IGFs) are associated with an increased risk of common cancers such as breast-, colon-, and prostate cancer.⁸⁵⁻⁹² These potent mitogens do not only stimulate cell proliferation and inhibit apoptosis, but they also interact with other molecules involved in cancer initiation and progression.⁹³ Some studies have shown that the level of IGF-I, one of the most important peptide hormones for growth and development, is influenced by other factors such as nutrition and physical activity – which suggests that lifestyle could play an important role in the regulation of naturally occurring mitogens.⁹⁴⁻⁹⁶

Lastly, the physical activity/anti-oxidant hypothesis is still in the early forms of development. DNA damage and repair occurs all the time. Every repair increases the risk of a misrepair, which in turn increases the risk of DNA restructuring, resulting in carcinogenesis.^{73,97} While a reduction of oxidative damage to DNA as a mechanism for cancer prevention is biologically plausible, studies have shown that DNA damage peaks 24–48 hours post exercise.^{97,98} Nevertheless, well trained individuals seem to have less oxidative damage after exercise, compared to unfit individuals and weekend warriors (people who only exercise once a week)^{99,100} but it is still unclear if physical activity inhibits cancer development in this manner.

Breast cancer

Breast cancer is by far the most common cancer among women worldwide^{101,102} with about one million new cases annually.¹⁰³ Incidence rates are high in developed countries such as Sweden. Almost 7,000 Swedish women annually, or 15–20 women daily, are diagnosed with breast cancer.¹⁰⁴ In less developed countries, the incidence rates are lower,¹⁰⁵ but expected to increase.¹⁰⁶

Women emigrating from countries with low breast cancer rates, experience increased rates when they move to high risk countries.^{107,108} The risk continues to rise for several generations until it eventually becomes similar to the rest of the population.¹⁰⁹ These findings have provided new insights into the etiology of breast cancer. While a number of studies already have shown that the hormone estradiol has a key role in the aetiology of breast cancer, the migration studies have revealed that exposure to a Western lifestyle has a substantial impact on breast cancer risk as well.

The concept of the Western lifestyle encompasses many factors and correlated exposures easily conceal true associations. Although epidemiology is the tool box to disentangle the relationships between risk factors and disease, this has been a challenge. Among other factors, obesity, alcohol consumption and physical activity have been on the agenda. They all have the wonderful appeal of being modifiable factors and thus realistic targets for primary prevention.

The effect of physical activity on reducing breast cancer risk is intriguing and has resulted in numerous studies conducted. But despite well-designed studies there is

inconsistency in the observed associations. The suggestive finding that physical activity in adolescence and young adulthood could protect against the later development of breast cancer deserved further study and was investigated as a part of this thesis.

Other site-specific cancers

Regardless of the different assessment methods of physical activity, or diverse populations, physical activity has been shown to have a significant inverse relationship with colon cancer risk.¹¹⁰⁻¹¹² The level of risk reduction is consistently around 40%.⁶ Many investigations included rectal cancer as a separate outcome. While the protective effect is strong and consistent for colon cancer, the risk of rectal cancer seems unrelated to physical activity.^{113,114}

The limited number of studies on physical activity and endometrial cancer has suggested a 20–40% decreased risk for women engaging in the highest levels of physical activity.^{6,115-117} The literature is still too limited and inconsistent to draw any firm conclusions about the relationship between physical activity and other cancer sites, such as prostate^{73,118,119} and ovarian cancer.^{120,121}

PHYSICAL ACTIVITY AND MORTALITY

Physical activity, weight control and the avoidance of tobacco smoking are three of the best documented modifiable factors that can prevent premature mortality in the general population.¹²² Several important investigations reported an inverse association between physical activity and overall mortality, cardiovascular disease mortality or mortality from other chronic diseases.^{37,38,123-127} Many of the earlier studies were done in men, and although the physiological response to physical activity ought to be equally beneficial for both men and women, the pattern of physical activity at work and at home may differ by gender. However, later studies established that physical activity reduces mortality among women as well.^{38,127-130} Despite promising research in the field, some issues had not been resolved. It was still unclear how cessation or changes in physical activity level from adolescence to adulthood affected mortality. In addition, it had not yet been convincingly documented whether increased physical activity was also effective among women aged less than 60 years.

We set out to study these knowledge gaps in a large cohort and calculate the fraction of mortality in the population of women below the age of 60 that could be attributed to physical inactivity.

5. AIMS

The overall objective of this thesis was to increase our understanding of physical activity as an exposure and the various methods of measuring it in epidemiological studies.

The aims of the studies included in this thesis were:

- To obtain a precise estimate of the association between physical activity during young adulthood and breast cancer risk using published data
- To assess the association between physical activity and all cause mortality among women below the age of 60
- To develop a novel instrument for self-reported total physical activity suitable for use in epidemiological studies
- To evaluate the validity and reproducibility of the new instrument separately for
 - work intensity alone and
 - usual physical activity
- To compare physical activity exposure patterns ascertained from questions commonly used in epidemiological studies with those obtained from the novel and validated instrument

6. SUBJECTS AND METHODS

The five studies making up this thesis differ in design, setting, size and methodology. However, they all tackle the challenge of measuring physical activity.

Table 2. Overview of the five studies included in the thesis.

Study	I	II	III	IV	V
Type of Study	Quantitative Review	Cohort	Cross Sectional	Cohort	Cross Sectional
Population					
Number of subjects	28,079	99,099	a) 94 b) 80	133+160	42,169
Age	20–80	34–49	16–65	20–59	18–94
Gender, Female	100 %	100 %	a) 40 % b) 61 %	53.2 %	64.3 %
Physical Activity Methods	Questionnaires with diverse methods	Questionnaire with 5-point-scale	a) Visual Analogue Scales b) Work on bicycle ergometer	Questionnaire with the instrument, 24-hour recall interview and test-retest	Questionnaire with the instrument and other commonly used methods
Outcome Measures	Relative Risk	Relative Risk	Correlation	Correlation	Correlation

STUDY DESIGNS, POPULATIONS AND PHYSICAL ACTIVITY ASSESSMENTS

Study I – Physical activity in a meta-analysis

Research synthesis – a background

In the era of information, the creation, distribution and manipulation of information has become an activity affecting the entire society. The medical research database has also seen rapid growth in the global information community. This makes the research synthesis an appealing and efficient approach to cover a topic (see figure 3).

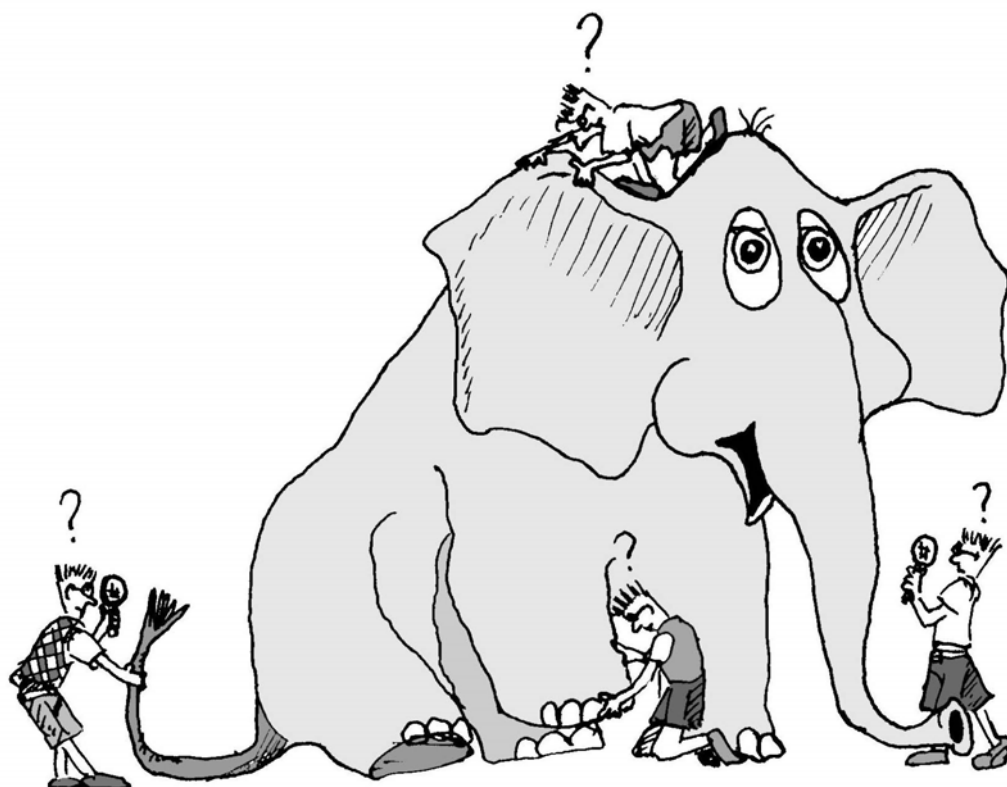


Figure 3. A research synthesis is an efficient way to cover a large topic. Drawing made by Henrik Trolle.

A research synthesis could be conducted in two ways. The classic narrative review is qualitative. It is often based on a selective inclusion of studies and by counting the assembled studies supporting various sides of an issue, the view receiving the most votes is chosen. This makes it prone to bias. On the other hand, the systematic quantitative review sets out to identify all relevant and valid pieces of information in the literature and does not ignore sample size, effect size or study design. These are important variables in the statistical analysis. This increases its ability to make an objective appraisal of the evidence.¹³¹

The term meta is Greek and implies something occurring later in time, beyond, transcending, more comprehensive.¹³² The term meta-analysis was first used 30 years ago and referred to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.”¹³³ Thus, the meta-analysis is the actual statistical analysis in the quantitative research synthesis and it does not refer to the entire project.

The meta-analysis can be seen as the epidemiology of results – where the finding of each study replaces the individual as the unit of analysis.¹³⁴ The research process does not differ from any other research undertaking in terms of a formulation of the problem to be addressed, collection and analysis of data and reporting of results.¹³⁵ The benefits of meta-analyses are multi-fold. It is a way to increase statistical power, resolve uncertainty when reports disagree, improve estimates of effect size, answer new questions not posed at the start of each individual study, improve the quality of primary research and give a more objective summary of the literature.¹³⁶

A research synthesis on physical activity and breast cancer risk

We were intrigued by the fact that findings regarding physical activity and breast cancer risk appeared to be too conflicting and confusing to allow causal inferences. Epidemiologists could not supply public health professionals with a clear message.

We chose to conduct a quantitative review to investigate whether vigorous physical activity during adolescence – the period of menses and breast development when physical activity is hypothesized to have a stronger protective effect¹³⁷ – was inversely associated with adult breast cancer. We also attempted to evaluate whether there was a dose-response relationship between physical activity during the same time period and risk of breast cancer.

We prepared a detailed written protocol in advance, with a priori definition of eligibility criteria for studies to be included. (See study I.) As recommended by Stroup *et al.*,¹³⁸ we used the following strategy to find and select relevant studies. Two independent reviewers identified original contributions of studies in humans, published since January 1966 and available on Medline. Besides the MeSH terms “exercise” and “breast neoplasm” we used “physical activity” and “breast cancer” for our search. Reference lists of identified articles and related reviews were also examined. Contact was initiated with authors to reduce the potential of publication bias and data sets used twice. Additionally, to find yet unpublished studies, meeting abstracts on breast cancer and physical activity were identified through the Web of Science.

When we initially conducted this search in May 1999 we found eight case control and three cohort studies. Three and a half years later, more original research, classic reviews and editorial comments had been published. They still disagreed on the issue. We now identified more than twice as many eligible reports. Included in our study were nineteen case-control studies and four cohort studies.

Methods of ascertainment of physical activity in adolescence and/or young adulthood were almost unique to each study, but all pertained to leisure time activity. We only included studies where a quantitative description of the physical activity variable was stated. Included in our analysis were studies measuring physical activity in h/week, times/week, times/year, METh/day/year, METh/week, METh/week/year, MET/week, MET scores and kcal/week.

Study II – Measuring physical activity the rudimentary way

As seen in the quantitative review, physical activity can be measured in a variety of ways. In the cohort study below we used a short single-item question to assess total physical activity.

The population-based cohort (The Women's Lifestyle and Health Cohort) was established in 1991 and 1992. In Sweden a sample of almost 100,000 women, born between 1943 and 1962 (aged 30–49 years) was randomly selected from the Swedish Population Register at Statistics Sweden (the council for official statistics). The source population was all women living in Uppsala Health Care Region (comprising about one-sixth of the Swedish population).

In Norway, a sample of 100,000 women born between 1943 and 1957 (aged 34–49 years) was randomly selected from the Norwegian Population Register. In this cohort the source population was the entire country at this time; women were randomly selected from the population from four five-year birth cohorts (i.e., age 30–34, 35–39, 40–44, and 45–49 years).

All women in the two countries received a letter of invitation and a questionnaire with, to a large extent, identical questions. In total, 106,841 women (54.5 %) returned the questionnaire. The study base consisted of 57,582 Norwegian women and 49,259 Swedish women.

From the initial cohort we excluded participants with missing vital status information and missing physical activity variables. Participants with a self-reported history of cardiovascular disease (myocardial infarction and stroke) and participants with a prior diagnosis of tumours from the cancer registries (except benign tumours, non-melanoma skin cancer, and non-invasive cervical cancers which we deemed unrelated to mortality) were also excluded. The final analysis was based on data from 99,099 women.

Using the national registration numbers, linkage to the Registry of Population and Population Changes made follow-up virtually complete. The women in the cohort were prospectively followed with regard to vital status and emigration through year 2003. The follow-up for each participant began the date of the return of the questionnaire and ended on the date of death, emigration or end of observation period, whichever came first.

Women rated their overall level of physical activity (i.e., physical activity in the household, occupational and recreational physical activity) at three time points: at age

14 and at age 30 retrospectively, and at enrollment. In the Swedish questionnaire, women ranked their level of physical activity on a 5-point scale with examples attached to level 1, 3 and 5. 1 = sedentary (mainly sitting), 3 = moderate physical activity (a few walks a week) and 5 = vigorous physical activity (sports/jogging several times a week). In Norway, women ranked activity on a scale of 1 (sedentary) to 10 (vigorous). We collapsed the 10-level Norwegian scale into five levels to be comparable to the Swedish scale.

Using the self-reported data at age 14, 30, and enrollment, we further categorized individuals on changes in physical activity over time. In order to preserve precision, we dichotomized physical activity exposure into women who participated in no or low physical activity (inactive), and those who participated in moderate, high or vigorous activity (active) for each time point. This categorization was based on similar risks of mortality in the groups. We then compared physical activity levels between age 14 and 30, age 14 and enrollment, and between age 30 and enrollment. From this comparison, women could be categorized as those who remained inactive, women who were active and became inactive, women who were inactive and became active, and those who remained active over the corresponding time periods.

Study III – Development of an instrument, a first validation study

Development of an instrument

The two aforementioned studies made it obvious that improved, validated and reliable methods to measure total physical activity in epidemiological studies were needed. We then set out to design such an instrument for self-reported total physical activity. The purpose was to develop a method that should capture all activities during every day life, regardless if they were undertaken as leisure time, occupational, or household activity. Furthermore, simple activities falling in-between these categories were not to be lost – optimally the instrument should be sensitive to these parts of daily life that are easily forgotten and often unfairly treated by epidemiologists. Last, but not least, it should record inactivity. Because all 24 hours per day are filled with physical activity or inactivity, all 24 hours were of interest.

Finally, we had one more objective; the instrument should allow estimation of total energy expenditure. Accordingly, the instrument had to be able to assess self-reported time spent on different absolute intensity levels during, for example, a typical day.

To achieve this, we developed a nine-step scale; each step was assigned a fixed value based on multiples of Metabolic Energy Turnover (MET). A one-step-increase corresponded to an increase in energy expenditure of $3.5 \text{ ml O}_2 \times \text{kg}^{-1} \times \text{min}^{-1}$ or $1 \text{ kcal} \times \text{kg}^{-1} \text{ body weight} \times \text{hour}^{-1}$ (1 MET). To increase the resolution at the sedentary end of the scale, we inserted half a step between 1 and 2 METs.

Contrary to, for example, the Borg scales, we anchored each step, not only the extremes. By anchoring in generally understood examples of activities, the instrument provided levels of absolute intensity, as opposed to relative, thereby enabling future studies to produce numerical data and translate it into an absolute measure of intensity on an interval scale level.

For each scale step we chose 2–4 example activities. To find suitable example activities which people perceive similarly in regard to absolute intensity, common every-day chores, job-related activities and sports were selected from the Compendium of Physical Activities,¹⁰ and translated into Swedish.

A first pilot study among six female and six male university employees revealed that some examples were confusing and needed elaboration. Riding became horseback riding for example. Next the activities were presented in alphabetical order to 94 adult volunteers, representing a wide range of education levels (included in the sample were 37 male and 38 female hospital staff members and 19 male construction workers).

In order to assess each subject's perception of the intensity of the activity, we used visual analogue scales (VAS). This method is particularly suitable in evaluating relative positions on a continuum. VAS is considered the "gold standard" for assessment of pain,¹³⁹ but it has also been used in appetite research¹⁴⁰ and to rate subjective feelings of exertion.¹⁴¹

VAS is typically composed of lines (usually 10 cm). The ends are labeled with the two extremes; the absolute minimum and absolute maximum that could ever be experienced. We used "No physical strain" and "The most physically demanding you could think of". See figure 4. These verbal descriptions anchor the ends, as the interpretations of extremes have the smallest inter-individual variation. Any intermediate point is non-verbal. Subjects are asked to make a mark somewhere along the line corresponding to their feelings. The marks are given a score between 1–100, by measuring from the left line to the mark with a ruler or a millimeter grid.¹⁴²

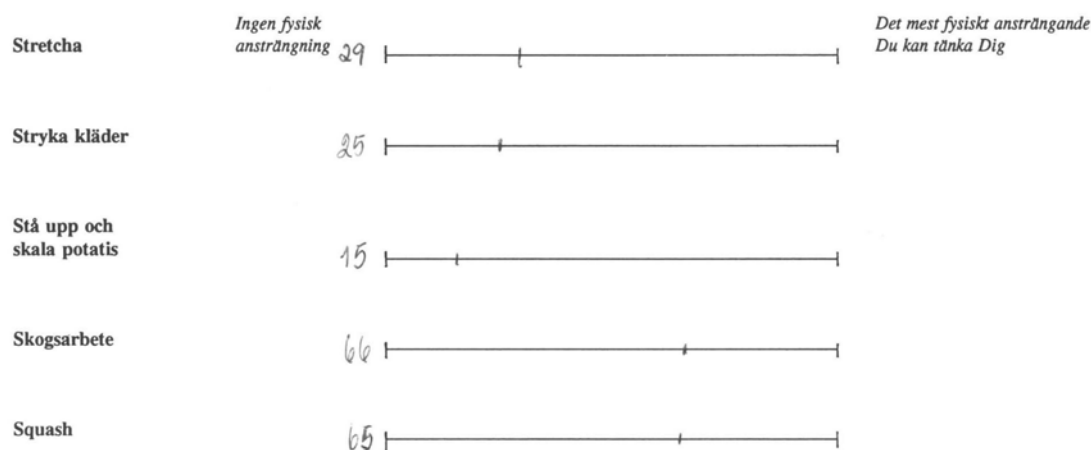


Figure 4. Visual Analogue Scales filled out by one of the subjects in the study. The scores, measured by the investigator in millimeters from the left mark, are also seen. The examples are stretching, ironing, standing and peeling potatoes, lumbering and playing squash.

The example activities chosen had to meet the following ordered requirements:

- 1) the rate of energy expenditure associated with an example activity had been measured objectively^{10,11}
- 2) most people should have a clear perception of the intensity of the activity
- 3) the perceived intensity should, on average, correspond to the measured absolute level of intensity. This requirement was met if the rank order of reported intensity corresponded with the rank order of absolute intensity derived from published rates of energy expenditure.
- 4) given that the previous requirements were met, the example activities to be selected should be the ones with the smallest inter-individual variation in people's estimation of the intensity.

As expected, the alphabetical list contained some examples of activities which most people in the general population have not and will not participate in during their lifetime. Despite the fact that not everybody might have painted a house, most have a clear perception of its absolute work intensity. Painting the outside of the house met the above requirements and serves as one of the examples for the 5 MET level.

We added two extreme levels to the final instrument; sleep/complete rest and activities more physically demanding than those exemplified in the 6 MET scale step. This was done to accommodate all possible subjective perceptions of extreme work load in order to ensure that such activities were not left out in the self-response, and to allow people to freely choose a level without the artificial constraints imposed by an instrument restricted to 1–6 METs.

A first validation study

Initially, we wanted to know if it was possible to accurately self-estimate concurrent work intensity on an interval scale level. We were not interested in any other aspect of the process of capturing physical activity in an epidemiological study. The ability of people to integrate their perception of physical activity over time to produce an estimate of the “usual” level and the accuracy of recalled activity are both critical for the successful evaluation of physical activity as a risk factor in epidemiological studies. However, first and foremost, we wanted to test the instrument against objectively measured work. This has, to our knowledge, not been done before in the development of a physical activity instrument for epidemiological studies.

We chose the bicycle ergometer to be our “gold standard”. The bike was the ideal choice for a gold standard because it has a direct relationship between physical performance and work load, it is easy to vary in a controlled way and it was unrelated to the activities in the instrument. Eighty participants (49 women and 31 men) between the ages 16–65 were recruited through an ad in the local paper. Each one was randomly assigned to four work loads corresponding to four out of the nine levels in the instrument. The bicycle ergometer was set for the work load/resistance level before the participant started cycling. Immediately after each cycling session the participant reported the perceived intensity using the instrument. This has been described in detail in paper III.

In order to evaluate reproducibility, twenty of the participants returned six weeks later to repeat the test in exactly the same manner as initially assigned – the same four work loads in the same order. The participants were unaware that they were completing the exact test as six weeks prior. Once again the participants reported the perceived work loads using the instrument.

Study IV – A second validation study in a random sample

Our next goal after we determined that the perceived intensity level attached to each step in the new instrument corresponded well with the true concurrent work load was to evaluate the instrument's validity and reliability in estimating "usual" energy. The highly complex cognitive task of recalling and integrating physical activity to produce an estimate of "usual" energy added a new dimension to our instrument.

The aim of this investigation was two-fold. In addition to evaluating the validity and reproducibility of self-reported energy expenditure using our newly developed instrument, we wanted to compare web surveys with traditional paper surveys in a cross-sectional design.

There are many advantages gained by computer-supported data collection that are not available with printed questionnaires. Above and beyond the logical benefits including a reduced cost per person, instant electronic storage, immediate checks for incomplete answers, and automatic summation, there are also other advantages such as inclusion of illustrations or sounds to clarify complex questions.^{143,144}

As this investigation was to be undertaken in a random sample of the general population, non-participation bias could potentially arise if inexperienced computer users chose not to participate when randomized to the web-based arm of the study.

Eight participants with no computer experience, three men and five women aged 55–67, were recruited for a qualitative pilot study. We aimed to investigate potential obstacles they would come across when, with the help of only written instructions, they were to open a web reader, go to the home page of the questionnaire and complete the survey.

The results of this study led to changes in the instructions, as well as in the layout of the web questionnaire. In the written instructions sent to the participants randomized to the web questionnaire we added pictures of the icons for web browsers, instructions on how to double-click on the mouse, scroll, and erase the old address (URL, Uniform Resource Locator) in the address bar. We also included a drawing of the keyboard showing where to find the erase button, shift and enter. Enter is needed to reload after writing the URL and shift is needed for the use of capital letters when entering the password. We learned that URLs with a forward slash (/) should be avoided, since they are particularly difficult to handle. In the final version we also changed the layout of the questionnaire to avoid unnecessary scrolling.

A second validation study

The study base was comprised of all subjects age 20–59, living in a typical middle-sized county in the northern part of Sweden. Subjects were randomly selected from the Swedish Population Register – a register updated yearly with 100 % coverage of all individuals living in Sweden. A sample of 765 men and women were invited to participate in the study and randomly assigned to receive either the traditional printed questionnaire ($n = 246$) or the web questionnaire ($n = 519$). Fifty per cent of each group was randomly assigned to the validation assessment of the new instrument and the other 50 % to the reliability assessment. For further details, see paper IV.

Three unannounced 24-hour recall interviews administered via telephone served as the gold standard. Results from earlier studies ¹⁴⁵ have shown that three 24-hour recalls provide an assessment of physical activity behavior comparable to other short-term physical activity assessments that utilized activity monitors.

Reproducibility was assessed by administering the instrument another three times during a period of six months.

Study V – Using the novel instrument in a cohort study

In this last study we set out to compare the physical activity pattern ascertained with conventional questions commonly used in epidemiological studies to that obtained using our novel and validated instrument. We had the opportunity to do so in a prospective cohort especially assembled to study the associations of energy expenditure through physical activity, dietary energy intake and anthropometric measures with various health outcomes.

Every second year the Swedish Cancer Society arranged a national four-day fund raising event, targeting not only schools and work places, but the entire Swedish population. With 38,000 volunteers from all over Sweden, every community had their own local festivities for cancer research. As part of these activities, local members of the community could buy tickets for a fund-raising walk. There were about 5,000 starting points for the walk, and the distances varied between 1–10 km. The payment for the walk was a donation of SEK 50 (one hour's pay at minimum wage at that time), to cancer research. This walk, called the National March, was a country-wide competition. The community with the highest participation rate won a visit by the Swedish King and Queen. The entire four day event, including fundraising galas and scientific programs on television, was heavily covered by the local and national media. See figure 5.

In 1997, the fourth time of the National March, we had the opportunity to reach this presumably motivated population by appealing for another aspect of personal donation – the donation of time. We asked volunteers to complete a 32 page color questionnaire, called “An Hour for Research” (see figure 6) covering thirteen different areas of lifestyle and health, with a special emphasis on physical activity. The participants were asked to provide their full national registration number, the unique personal identifier assigned to all Swedish residents. This would allow multiple record linkages with existing nationwide, continuously updated and essentially complete databases such as

the population register and the registries for migrations, causes of death, cancer, and hospital care.

The participants were invited to take a questionnaire at the starting point and bring it home to fill out at their leisure. To return it, a deal was made with one of the large chains of super markets across Sweden. Their 2,300 stores were all equipped with a special mailbox, making it possible for the participant to securely post the questionnaire at the local super market, no postage required. Next, the questionnaires were collected when the milk trucks exchanged milk for mail boxes. The mail boxes were then transported free of charge to Statistics Sweden for scanning. Participants could, if they wanted, send the questionnaires by regular mail, but due to lack of funding, they had to cover the postage themselves. In total 43,880 participants completed and returned the questionnaire.



Figure 5. The launch of the National March Cohort was covered by the media.

The questionnaire covered thirteen different areas of interest, but started with several pages inquiring about physical activity patterns. We used the abovementioned instrument in addition to several other conventional ways to measure physical activity level in epidemiological studies.

Participants were asked to indicate the average number of hours per week they participated in leisure time physical activities such as sports, exercise or outdoor activities during summer and winter in three levels; light (exemplified by casual walking), moderate (exemplified by walking in a brisk speed, jogging and swimming) and heavy (vigorous training or competition). Additionally, we inquired about the average number of hours spent a week on household exercise such as cleaning the house, working in the garden, and walking and/or biking to work. Each level was given a value in terms of multiples of Metabolic Energy Turnover (METs) based on the

estimated oxygen consumption needed for that level of activity. These were 3.0, 6.0 and 10.0 METs for the exercise questions and 4.0 METs for the question on household exercise. The MET-value was multiplied with the number of hours engaged in the particular activity level, resulting in an index of activity (energy per body mass) measured in MET_h. See figure 2. Furthermore, participants were asked to rank their own relative fitness and relative total physical activity compared to others of the same age and gender on a five-point scale, (much worse/less, worse/less, approximately the same, better/more, much better/more).

**Figure 6. The Swedish Queen with our questionnaire “An Hour for Research”.
Photo by Jacob Forsell.**

STATISTICAL METHODS

Study I

Conducting a meta-analysis

When conducting a quantitative review, there are two conventional statistical models to choose from, the fixed effects model and the random effects model. The key issue is homogeneity.

If we assume we have homogeneity across studies then we assume that the results of the studies we are to combine estimate the same true result and that the observed differences between them are due only to sampling variation. In addition, we also assume that true values do not differ from one situation to another.¹⁴⁶ In this case, a simple pooling of the results will suffice. For example, if the prevalence of a particular disease or characteristic is the same in each health care region in Sweden, then the national prevalence can easily be estimated by pooling the numbers. The prevalence is

then given by the total number of cases of the disease in the total population and can be reported as an overall proportion.

When we choose to use the fixed effects model, based on the assumption of homogeneity, we might not be able to honestly say that we believe all studies estimate exactly the same thing. However, we assume they are close enough; at least to the extent that the result will not mislead us.¹³⁶

If we instead believe that every study has its own truth, we use the random effects model which accounts for both within-study sampling variation and between-studies variability. This will result in wider confidence intervals than in the fixed effects model. However, if there is homogeneity in our data, the random effects model will be reduced to the fixed effects model.^{133,136}

The random effects model was the natural choice in our study of physical activity in pre and postmenopausal women. We combined studies conducted in different countries, with different designs, methods and units of measurements of physical activity. We conducted a summary analysis on the relative risk (RR) comparing the highest to lowest category of physical activity, regardless of the method used in the original study.

We statistically examined the degree of similarity in the outcomes of the studies. Unfortunately, the tests lack power – they often fail to reject the null hypothesis of homogeneous results even if there are large differences between studies.¹³⁵ But heterogeneity should be viewed as expected and not as an exception in the meta-analysis of epidemiologic data.¹⁴⁷ Even if we have non-significant heterogeneity, exploring, assessing and reporting causes of variation is reasonable and useful. Identifying sources of variability in results across studies can lead to valuable knowledge useful for planning and conducting future studies. We conducted subgroup analyses, hypothesizing that study location, study design, study size, intensity of physical activity, adjustment for potential confounders and menopausal status could explain inter-variability in the results. We conducted sensitivity analyses lest confounding factors not adjusted for in three of the studies influenced the final summary estimate.

Finally, we estimated the relative risk of breast cancer associated with each hour increase in weekly exercise with a mixed effect weighted regression model. In order to treat hours of exercise per week as a continuous exposure variable, its value was set at the midpoint of each category presented in the individual studies. For the open-ended highest category, such as “≥ X hours of exercise per week”, the approximate midpoint was set at double the distance between midpoint and upper bound of the closest category.

The file-drawer method and other ways to test for publication bias

No matter how thorough the search is for studies to include in a meta-analysis, the studies found will not reflect all the studies ever done. One reason could be that studies are tucked away in the file-drawers of researchers around the world, simply because

they do not make the magical 0.05 level – they are not significant and thus never submitted.

We can deal with this aspect of publication bias. The most well-known technique is Rosenthal's "file-drawer" method.^{133,148} The aim is to calculate how many new, filed or unretrieved studies with an average observed effect of zero are needed to bring down the result of the quantitative review to barely significant. There is no exact number for what is the likely number of unpublished studies; this depends on the area of research. However, if the number is reasonably high, one can conclude that the observed result of the quantitative review is unchallengeable and robust to potential publication bias.

We also made a funnel plot where we plotted the size of the effect against the standard error of the study. Furthermore, we conducted the rank correlation test,¹⁴⁹ the statistical test that is analogue to funnel plot and was created to provide quantitative evidence against the hypothesis that no publication bias exists.¹³³

Study II

Analyzing a cohort study

As an initial level of analysis of the relation between physical activity and total mortality, we calculated crude cumulative incidence (Kaplan-Meier) curves. The Kaplan-Meier method of estimating a survival function uses the exact survival times for each subject instead of grouping them into intervals.¹⁵⁰ The curves, based on one record from each of the 99,099 women in the cohort, gave us an unadjusted descriptive picture of what happened in the cohort during almost 12 years of follow-up.

We then proceeded with adjusted analyses in the Cox Proportional Hazards Model, which models the risk of the event (all cause mortality) up to each point in time. Low levels of physical activity were used as the reference. Initially, we did not control for any potential confounders. We estimated the crude hazard ratio (HR) and the 95 % confidence interval (CI). However, age is an important determinant of mortality and although there was little difference in physical activity intensity by age (maybe because we had a relatively narrow age range of women at baseline) one may well think physical activity actually diminishes with age (at the end of the study the highest age reached was 62 years). Thus, we chose to present age-adjusted HR, especially since it is possible to calculate crude rates and thus rate ratios from the tables.

Next we adjusted the estimated HRs for the following potential confounders: age at enrollment, years of education, body mass index, alcohol intake, smoking status, mean number of cigarettes smoked per day while smoking, number of years smoking, and country of origin. Each confounding variable was carefully chosen.

We chose age at enrollment as an adjustment variable for several reasons. Physical activity level was only measured at enrollment. Furthermore, in this paper, we were interested in the effect of physical activity, conditional on age (and the other covariates). Using age as a baseline covariate, the model would look within risk sets of those in the same age at study entry each time someone died. In this way, we could

control for age. We preferred to use age as a baseline covariate and length of follow-up as the time to event variable to track aging, since the two are related.

We used education as a proxy for socioeconomic status, which we know is a strong predictor of mortality in many settings. We do not know the exact mechanism for this, although education could be a surrogate for stress, poor diet, etc. In our data, education was a predictor of mortality, although only slightly associated with physical activity.

BMI did not follow a linear relationship with mortality. Death was high in the lowest group (< 18.5), lower for 18.5–24.9, and then it increased again, making a j-shaped curve. We were worried that if we modelled BMI as a continuous variable, we would “miss-specify” the form, and thus not control entirely for the confounding. We chose to make it a categorical variable.

Alcohol intake per day was also modeled as a categorical variable with 0, <1 drink, 1 or more drinks per day. According to IARC’s (the International Agency for Research on Cancer), advice for analyses, 1 glass of any drink contains approximately 10 g of ethanol no matter what the alcoholic beverage is.

Smoking was dealt with in several ways. Smoking status (categorical: current, former, never), mean number of cigarettes smoked per day while smoking (continuous) and number of years smoking (continuous) were included in the model.

Finally, we included country of origin in the model (categorical: Norway vs. Sweden).

In the third model we also included physical activity at each point in time to separate physical activity at baseline from past physical activity levels.

Some variables could conceivably be in the causal pathway between physical inactivity and an increased risk of death. To assess possible effect modification, we estimated HRs comparing active versus inactive women and stratified by age, body mass index, smoking status, education and country of origin. We also modelled interaction terms and calculated log-likelihood ratio tests to compare how well competing hypotheses were supported by the data.

To assess the impact of physical activity on mortality in the population, we calculated the population attributable fraction (PAF) under two assumptions: that each woman would switch to the next level of physical activity, except those already in the highest level of activity, and that women who were inactive would become active. To calculate the PAF we used estimates of the relative risk derived from our proportional hazard models, adjusting for confounders, using the formulae presented in Rockhill *et al.*,¹⁵¹ Bruzzi *et al.*,¹⁵² and Bosetti *et al.*,¹⁵³ $1 - \sum_i pd_i / RR_i$ where pd_i is the proportion of exposed cases in the i th category, and RR_i is the relative risk from the proportional hazard model comparing i th category to the reference group (we used “vigorous” as our reference $1.0 = 0.50 / 0.50$). See table 3 and the table found in the online supplement, paper II.

Women who were ill at the time of enrollment may have had a different physical activity level than women who were not ill, and may also have been at an increased risk of death. Thus, besides excluding women with an a priori diagnosis of cancer or cardiovascular disease before cohort enrollment, we conducted several sub analyses within the data to assess possible bias associated with pre-existing illness. First, we excluded all women who reported themselves as having fair or poor health at enrollment. Then, we excluded women who had a self-reported diagnosis of diabetes or hypertension. Next, we excluded deaths that occurred during the initial two or three year of follow-up. Since the results from the analyses on the impact of physical activity on overall mortality were virtually the same, with and without these exclusions, we chose to present results including all women in our cohort, except the women with a priori diagnoses of cancer or cardiovascular disease.

Study III

Correlation analyses in the first validation study

In the following three studies we were interested in quantifying the strength of the linear relationship between two variables. Since we wanted to estimate the linear correlation of the underlying continuous measurement of METs we deemed parametric statistical methods appropriate in study III.

We estimated a Pearson correlation coefficient between performed and reported intensity for all sessions on the bicycle ergometer ($n = 320$). This analysis considered all estimates to be independent. But since each subject contributed four pairs of intensities, the estimate could be influenced by within subject (dependent measure) confounding. To avoid this, we randomly selected only one pair of intensities for each subject and conducted the regression analysis on this set of 80 independent data points.

Since the novel instrument is based on steps, one could be deceived to believe the data is categorical and thus use nonparametric tests such as Spearman's rank correlation coefficient. However, we were not primarily interested in knowing if subjects were able to rank performed work loads correctly. The rank order of the four work loads could be perfect with a correlation of 1, despite the fact that a true 1 MET increase in performance might be perceived to be an increase of several METs.

We computed intraclass correlation coefficients to evaluate the test-retest reproducibility. This is the measure of reproducibility of repeat measures from the same subject. It varies between 0, no reproducibility, (i.e., large within-person variability and no between-person variability) and 1, perfect reproducibility (i.e., no within-person variability and large between-person variability).¹⁵⁴

Finally we also conducted multiple regression analyses to study whether background variables influenced the ability to correctly report work load.

Study IV

Correlation analyses and other approaches in the second validation study

Once again we wanted to estimate the correlation between two measurements. Basically we conducted the same analyses as in paper III, followed by some complementary approaches.

The correlation coefficient quantifies the degree to which the two measurements are linearly related. However, Pearson correlation coefficient fails to detect non-reproducibility when we, for example, have a scale shift or a location shift.¹⁵⁵ A high correlation does not necessarily mean that the two methods agree. We have perfect agreement if all the points lie along the line of equality, (the line all points would lie on if the two methods produced exactly the same measurements) but we have perfect correlation if the points lie along any straight line.¹⁵⁶

To better assess agreement between the two methods, Lin's concordance correlation coefficient was used to evaluate the degree to which pairs fall on a 45-degree line from the origin (or a slope of exactly 1.00).^{155,157} Lin's concordance correlation coefficient contains the measure of precision (as does the Pearson correlation coefficient) but also the measurement of accuracy. Thus, even if precision is 1, any departure from the line would produce a correlation coefficient of <1.

We also constructed a Bland-Altman plot. By plotting the difference between the measurements of the two methods for each subject, against their mean, one can better investigate any possible relationship between the measurement error and the true value. However, since we do not know the true value, the mean of the two measurements is the best estimate we have. We can not use the measurement of each method separately, because the difference is related to each.^{156,158}

Any source of variability not due to true differences between subjects contributes to total variance and reduced reliability. Furthermore, a large variance may change the mean, or mask correlations and bias towards the null.¹⁵⁹

In our study, variance could arise from two major sources: inter-individual variation (the true between-person variation in physical activity habits) and intra-individual variation, including true day-to-day variation within subjects and methodological errors. To correct for random within- and between person variability we used a method described in the diet assessment literature.¹⁵⁹

We used the one-way random effects model ANOVA to estimate the within and between person variance to evaluate test-retest reproducibility.

Finally, we conducted linear regression modeling to evaluate potential differences in the ability to self-report energy expenditure across respondent characteristics.

Study V

Cross-sectional analyses in the novel cohort

The 110 shelf meters large National March Cohort data was treated like a brandy, left to mature in casks, waiting for the increased value that comes with time – to become very special (VS). And while the brandy is still maturing for future research, we, as any brandy maker, tested the quality at an early stage. This was done by conducting cross-sectional analyses.

First, we linked the entire cohort to Statistics Sweden's Total Population Register to check the validity of the national registration numbers. Subjects giving invalid, incomplete or no national registration number cannot be followed up with record linkages. Seventy eight national registration numbers could not be found in the Total Population Register. They will be excluded in future analyses as loss to follow up.

We were interested in the characteristics of this novel cohort. As the questionnaire was rather comprehensive, we had 440 variables to choose from. We chose to look at age, gender, BMI, cigarette smoking, alcohol consumption, education and different aspects of physical activity with simple tabulations. We manually checked the answers where the scanning had not worked properly. Moreover, we manually checked potentially unreasonable answers such as weight and length resulting in a BMI <15 or >50 kg / m² or males having answered questions in the female section. The data quality was in general very good.

With the aim to describe concordance and disagreement between different measures of physical activity, we conducted Spearman's rank correlation analyses. With perfect correlation between two variables, the ranks for each variable would be the same. We chose a non-parametric statistical method as some variables were ordinal and the underlying distributions were not assumed to be normally distributed.

We moreover conducted linear regression modeling to study the independent effect of gender, age, BMI and education on total physical activity measured with the new instrument and leisure time physical activity measured with conventional questions.

7. RESULTS AND COMMENTS

STUDY I

Physical activity in adolescence and young adulthood and breast cancer risk: a quantitative review

Our summary analysis revealed a significant inverse association between moderate to vigorous physical activity during adolescence/young adulthood and breast cancer. The reported relative risk (RR) of high versus low physical activity, ranged from 0.2 to 1.4 in the nineteen case-control and the four cohort studies included in our analysis. See figure 1, paper I. The summary RR for all the studies was 0.81 (95 % CI 0.73–0.89). The inverse association between physical activity and breast cancer was consistent despite variation in populations and methods, but we found significant heterogeneity across all our results. The summary RR estimate for each 1 hour/week increase in exercise was 0.97 (95 % CI 0.94–1.0). See figure 7.

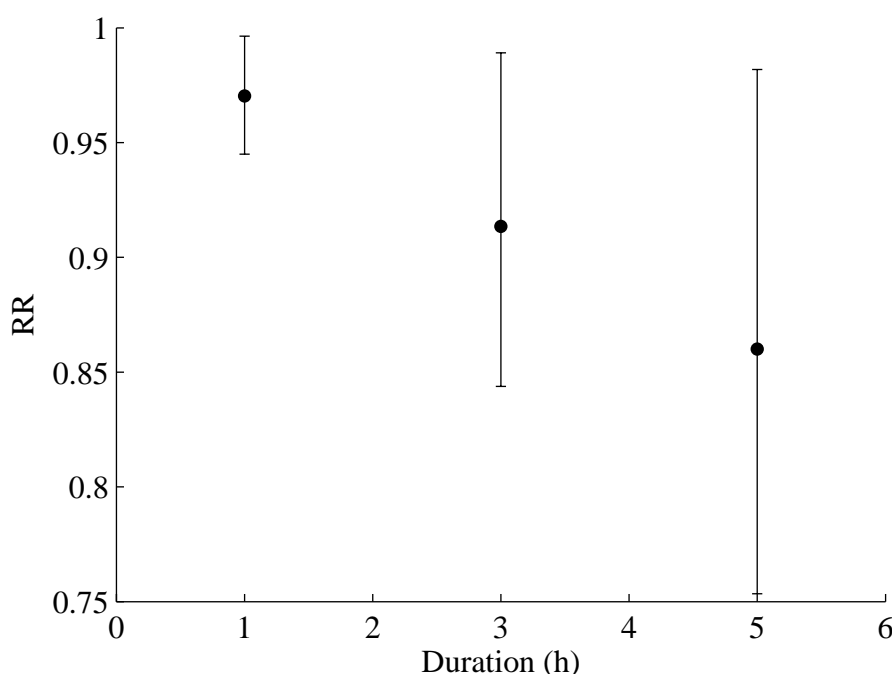


Figure 7. The relative risk (RR) and 95 % confidence interval for weekly hours of exercise at leisure time during adolescence/young adulthood.

Originally, when we conducted the meta-analysis on 11 studies, the distortion (“file drawer”) analysis showed that the results were not resistant to publication bias. With the addition of just one single non-significant study on exercise and breast cancer risk, the hypothetical pooled estimates of the relative risk in the highest vs. the lowest analysis were no longer statistically significant. This was a discouraging result, one single unpublished study – well, between us, the three coauthors, this could be found in our drawers! However, as the data base grew and we could conduct our meta-analysis on 23 studies, the significance of the observed effects became robust to potential publication bias. Now we would need to find 134 or more non-significant studies on exercise and breast cancer risk for the relative risk to no longer be significant.

Furthermore, even though a low powered test, the rank correlation test for publication bias was not significant either.

Publication bias could be suspected in our quantitative analysis, since when stratifying by study size, the smallest studies showed the largest effects. The funnel plot could be seen as a diagnostic plot to detect publication bias. Without publication bias the plot should be “funnel” shaped with a wider spread among the smaller studies making up the base of the funnel, and a smaller spread among the larger studies, making up “the neck”.¹⁶⁰ See figure 8.

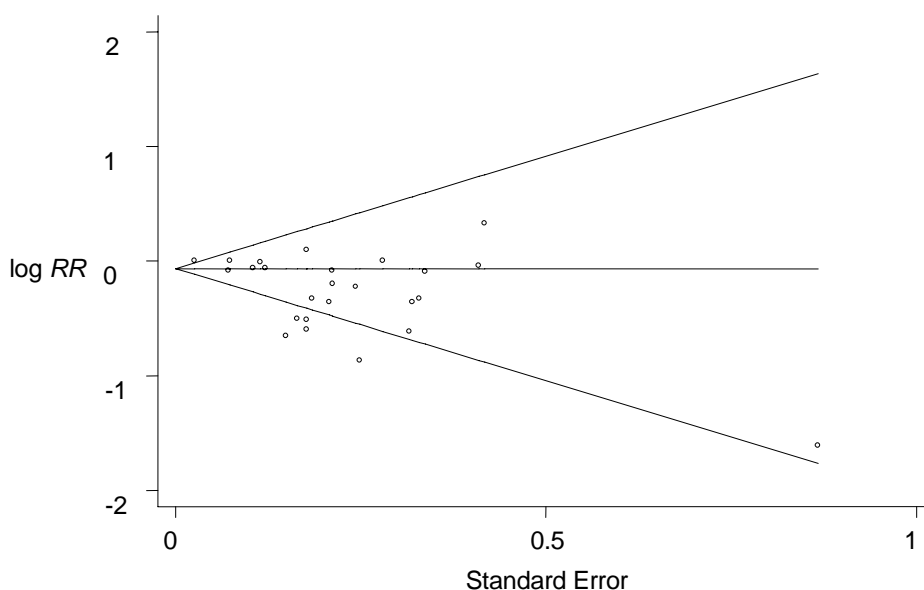


Figure 8. Begg’s Funnel Plot of the 23 studies included in study I, with pseudo 95 % confidence limits. With no heterogeneity the studies are spread uniformly on both sides of the summary estimate. Note that log RR for no effect is zero (corresponding to a relative risk of 1).

STUDY II

Physical activity as a determinant of mortality in women

In this large, population-based, prospective investigation among women, we found that physical activity significantly and substantially reduced overall mortality. Ten years after enrollment, the cumulative mortality was 1.55 % (95 % CI 1.37–1.72) among women with no or low activity, compared with 0.87 % (95 % CI 0.76–0.98) among those with high or vigorous activity.

Risk of death decreased steadily over five categories of physical activity at cohort enrollment (p for trend <0.0001). Compared with women in the lowest category of physical activity at enrollment, those in the highest category had half the risk of dying (HR= 0.46; 95 % CI 0.33–0.65) after adjustment for potential confounders. In contrast, there was no convincing dose-response trend, neither for physical activity at age 14 nor at age 30. When adjusting for current physical activity, the associations with physical activity at age 14 and 30 disappeared. This observation suggests that there is no direct effect of physical activity early in life on reductions in total mortality.

In table 3, women are distributed according to whether they changed or remained in the same physical activity category over time. These data show that physical activity levels in the past are strongly related to activity levels in the future.

Table 3. Changes in physical activity over time.

	Changes in Physical Activity Level (%)			
	Inactive-no change	Active to Inactive	Inactive to Active	Active-no change
At age 14 to age 30	4.3	9.9	8.2	77.6
At age 14 to enrollment	3.7	16.9	8.9	70.5
At age 30 to enrollment	9.2	11.4	4.9	74.5

Inactive is defined as participating in no or low physical activity.

Active is defined as participating in moderate, high or vigorous physical activity.

The effect of physical activity at enrollment on overall mortality was consistent across age, history of smoking, educational level, and country of origin. There was, however, evidence of statistically significant interaction of physical activity with body mass index. The effect of physical activity was more accentuated among women with a BMI <18.5 (see table 4).

Table 4. Modification of the effect of physical activity at enrollment on risk of death by age, smoking status, body mass index, education and country of origin by changes in physical activity level, 1991–2003.

	None/low	Moderate-vigorous	P for interaction
Age Group at Enrollment			
30–39	REF	0.72 (0.56–0.94)	0.79
40–49	REF	0.70 (0.60–0.82)	
Smoking Status			
Current	REF	0.65 (0.54–0.79)	0.82
Former	REF	0.75 (0.60–0.93)	
Never	REF	0.71 (0.54–0.92)	
Body Mass Index			
<18.5	REF	0.32 (0.16–0.63)	0.029
18.5–24.9	REF	0.77 (0.65–0.91)	
25.0+	REF	0.59 (0.48–0.73)	
Body Mass Index			
<25.0	REF	0.73 (0.62–0.86)	0.56
25.0+	REF	0.59 (0.48–0.73)	
Education			
<12 years	REF	0.69 (0.59–0.82)	0.97
12 or more years	REF	0.70 (0.58–0.85)	
Country of origin			
Norway	REF	0.74 (0.62–0.87)	0.27
Sweden	REF	0.63 (0.52–0.78)	

Data are adjusted for age at enrollment, years of education, body mass index, alcohol intake, smoking status, mean number of cigarettes, number of years smoking, and country of origin.

We examined this further and table 5 presents the overall cumulative incidence of mortality by BMI and physical activity level. Among those with the leanest BMI (<18.5), mortality was higher among those with no/low activity, than for those who were physically active. Interestingly, mortality was elevated regardless of physical activity level among women with a BMI of 30 or more.

Animal models have shown that caloric restriction combined with exercise result in an increased average survival.¹⁶¹ In humans, it has been reported that unfit lean men have double the risk of all cause mortality compared to fit lean men,^{123,162} indicating that excess mortality in the leanest is also concentrated to the inactive. This is supported by our data on women, where the incidence of death is elevated for the leanest and inactive women.

Table 5. Cumulative incidence of mortality by physical activity at enrollment and BMI.

Physical activity at enrollment	Body Mass Index			
	<18.5	18.5–24.9	25.0–29.9	≥30.0
None / Low	4.23	1.62	1.79	3.79
High / Vigorous	0.94	1.02	1.13	2.22

Given the apparent protective effect of physical activity at baseline on mortality, we calculated the population attributable fraction (PAF) to quantify the overall reduction in mortality that would be experienced by the population. If women in Norway and Sweden were to become more physically active, defined as moderate to high physical activity, overall mortality in this population would be reduced by 9.2%. Alternatively, if all women were to switch to the immediately higher physical activity level, except for those already vigorously active, then overall mortality would be reduced by 11.9%. And finally, if all women would engage in vigorous physical activity, then mortality would be reduced by 23.9%.

STUDY III

Validity of perceived work intensity using a novel instrument

In our attempt to develop a new instrument for self-reported total physical activity in epidemiological studies, the first validation study gave evidence that under standardized conditions it was possible to obtain valid and reliable self-reports of concurrent intensity using our generic instrument with scale steps defined in terms of examples unrelated to the work being assessed.

A high percentage of reported intensity level, 91%, fell within ± 1 MET of performed intensity level, I_{true} . Underestimations of intensity were more common than overestimations. The proportion that underestimated I_{true} increased as the actual intensity of the work increased.

The Pearson correlation coefficient was 0.96 (range 0.73 – 1.00), assuming the 320 estimates were independent. The mean Pearson correlation coefficient from the 80 truly independent estimates was 0.89 (range 0.81 – 0.95). See figure 9. When we randomly chose one pair of intensities from each participant in the test-retest scenario, the intraclass correlation coefficient was 0.99 (95 % CI 0.98–1.00).

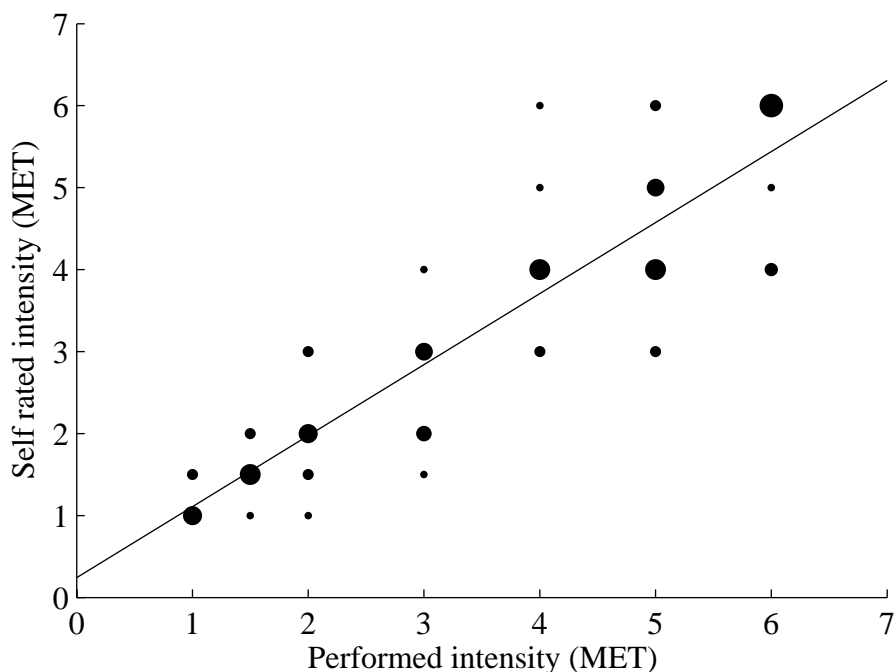


Figure 9. The concordance between the performed intensity level (I_{true}) and participants' reported intensity level (I_{reported}), both expressed in METs for one randomly chosen observation per participant ($n = 80$).

STUDY IV

Validity and reliability of self-reported total energy expenditure using a novel instrument

Our second validation study shows that it is possible to obtain valid and reproducible self-reports of usual physical activity, translated into total energy expenditure, using our novel instrument. The instrument is well suited for use in the general population and on both printed and electronic self-administered questionnaires.

Pearson correlation between usual daily energy expenditure measured by the instrument and our gold standard, the mean of three 24-hour recalls, was 0.73. On average, usual daily energy expenditure was 3,499 kcal, while the mean energy expenditure measured by the three 24-hour recalls was 2,819 kcal. There was a tendency towards overestimation of usual physical activity. This was significantly associated with low education. Reproducibility showed an intraclass correlation of 0.55. The study demonstrates that randomly selected individuals are able to convey a fairly good estimate of their total physical activity with the new instrument.

STUDY V

Assessments of physical activity in epidemiologic studies are sensitive to the method of inquiry

Our cross-sectional analysis demonstrates that the precision of self-reported physical activity depends on the manner in which the question is asked and the type of responses offered as options. It summits some of the complexity of the measurement of physical activity.

Total physical activity measured by the new instrument was underestimated, but corresponded to an average of 32.8 METh/day or 2,353 kcal/day. Household and leisure time physical activity comprised 17 % of total physical activity measured by the instrument. The Spearman correlation coefficient between the two measures, total physical activity measured by the instrument and household and leisure time physical activity, was 0.26.

The Spearman correlation between self-ranked relative total physical activity and quintiles of METh for total physical activity generated by the new instrument was 0.19, while the Spearman correlation between self-ranked relative fitness and quintiles of METh for total physical activity obtained with the new instrument was 0.20.

Body mass index ≥ 25 , education ≥ 12 years, and age ≥ 60 years were significantly associated with lower total physical activity as measured by the new instrument.

8. GENERAL DISCUSSION

METHODOLOGICAL CONSIDERATIONS

An epidemiologic study could largely be seen as an exercise in measurement.¹⁶³ We estimate exposure, we estimate outcome, and we estimate, for example, the health status or disease frequency of a population. This process of estimation ultimately comes down to the issue of measurement. We strive to minimize error in our measurements to be able to make correct estimations. But error can sneak into our studies. Error can be either random or systematic and it can occur before we even know it.

Validity

The five papers that make up this thesis mention the words validity, valid and validation 78 times. Validity is undoubtedly one of the most important issues, not only within this thesis, but also within epidemiology. Validity refers to two concepts; internal validity and external validity, also known as generalizability. Internal validity is both necessary and regarded as precedence over external validity.

With internal validity, i.e., accurate measurements of effects and no random errors, it is possible to draw inferences from the study to the source population – the cohort from which subjects are drawn. In study II, this would refer to Norwegian women and women in the Uppsala Health Care Region. External validity refers to the ability to draw inferences to people outside that population – to the target population. Our target population is women. Not just female Uppsala residents and Norwegian women, but women overall.

Bias

Bias is a systematic error that alters our estimates. In order to ensure validity, bias needs to be avoided. Although the same validity issues can be viewed from different angles and classified various ways, I choose here to present two main groups: selection and information bias – where the investigator is responsible, and confounding – where nature is responsible.

Selection bias

When the study population is not representative of the theoretical cohort of all eligible subjects, there is selection bias. This bias can occur either when there is a biased sampling of the population or a differential loss of subjects during follow-up.¹⁶⁴ Although selection bias can be prevented because it occurs in the study design, it can never be fully corrected by data manipulation after the study is complete. Even in well thought out studies, there is never an absolute guarantee that bias is eliminated. Selection bias was a concern in all our studies.

It should be emphasized that the National March Cohort, study V, does not constitute a representative sample of the population. As athletic associations and sports clubs were co-organizers of the event at which the cohort was recruited, it is reasonable to assume that physically active individuals were overrepresented. Moreover, the participants had

to be motivated and healthy enough to make it to their local event to pick up the questionnaire. Therefore, data on exposure prevalence and cross-sectional associations should not be uncritically generalized to the general population. A majority of the participants indicated that they were more physically active than their peers and when we explored the data, we saw that participants in the cohort smoked less than Swedes in general. As the social desirability of non-smoking and physical activity is indisputable, this could have led to a tendency towards overestimation of these behaviors.

The Women's Lifestyle and Health Cohort, study II, on the other hand, was based on a population sample and thus was not subject to the same problems as seen with self-referral. However, the response rate did not exceed 54.5 %. To investigate the potential risk of selection bias due to loss of follow-up, analyses of validity in the Norwegian part of the cohort were conducted. However, they revealed no major source of selection bias; there were no differences in life-style factors between the responders and non-responders.¹⁶⁵

Selection bias was also a concern in study IV, the second validation study. Complete data on only 38 % of the participants (293/765) were available. It is possible that we had both a biased selection into the cohort and a selective loss to follow-up. More than 50 % of our participants had completed >12 years of education. We investigated the frequency of higher education in the entire population of Sweden. During 2001–2003, when the study was conducted, 40.7 % of the population was educated for >12 years.¹⁶⁶ Thus, our participants were more educated than the average Swede. If higher education is associated with more valid answers, the overall performance of our instrument may be somewhat overestimated. However, one of the strengths of this study is that it was based on a representative sample of the general population. Thus, our responders are likely to be representative of the special types of people who are expected to participate in epidemiological studies that use similar recruitment approaches. In our assessment of response profiles all invited subjects, whether participants or not, were informative.

In spite of our attempts to select volunteers representing a wide variety of ages and fitness levels in study III, they did not constitute a representative sample of the population. Our participants may have been more motivated and apt to use self-rating instruments. But for practical and ethical reasons, the study was based on a healthy, self-referred sample which could have introduced a bias. For example, ill individuals may have perceived physical exertion differently.

And finally, in the meta-analysis in study I, we discuss and test for publication bias, which could be considered a form of selection bias, since only published studies provide the basis for analysis.

Information bias

Bias can also be introduced when obtaining information on exposure or outcome. Similar to selection bias, it occurs in the study design and can be prevented, but it can never be corrected by data manipulation once the study has been conducted. Information bias occurs when we do not measure the exposure and/or outcome well

enough and as a consequence classify the participants incorrectly.¹⁶⁴ This misclassification may be nondifferential or differential. Assume for each subject we measure exposure A and outcome B. Nondifferential misclassification occurs when the error in classifying subjects on exposure A is independent of the classification of outcome B and vice versa. However, if the degree of misclassification on exposure A varies according to the category of outcome B, or B varies according to the category of A, the classification is nondifferential.

This may not seem to be of major importance; however, this way to distinguish between the different types of information bias is valuable. Nondifferential misclassification leads predominantly “to bias towards the null”, i.e., to an underestimation of the association, the relative risk, between the exposure and the outcome. Although, in certain circumstances it could produce bias away from the null leading to higher estimates than the true ones.¹⁶³ Differential misclassification goes in either direction; it could either exaggerate or underestimate an effect.

Classification errors can occur in several stages of a study – from imprecise measurements of the exposure to imprecise self-reports to imprecise, missed or mistaken diagnoses/outcomes. Nondifferential misclassification is presumably present in every epidemiologic study to some degree¹⁶³ and consequently also in the ones presented in this thesis.

It is conceivable that we have misclassification of exposure in study II, the Women’s Lifestyle and Health Cohort. A crude assessment indicated the women's perception of their own physical activity at different stages in life and it has never been validated. This self-assessment can introduce misclassification, but we have no reason to believe this error is correlated with outcome. Outcome – death or survival – is measured essentially without error because of linkage to virtually complete nationwide population registers. Any misclassification of outcome would be of the nondifferential type.

Another type of misclassification is illustrated in study IV, the second validation study. Compared to the estimates obtained during an interview, self-reported usual physical activity was overestimated. Moreover, participants with lower education were more likely to overestimate their physical activity, suggesting differential misclassification.

When binary variables are classified (or misclassified), sensitivity and specificity are two measures of validity. This is most commonly used in screening, but could be calculated in a validation study as well. If we want to determine the misclassification rates of the self-reported measures compared to the “true” status obtained from medical records for example, sensitivity is defined as the probability that a person reported exposure, given the person was truly exposed. Specificity is defined as the probability that a person reported no exposure given the person was truly unexposed. This relates to predictive value. Predictive value positive is the probability that a person was truly exposed given that the person reported exposure, while predictive value negative is the probability that the person was truly unexposed given the person reported no exposure.

Reliability is the ability of a method or a test to give the same result when repeated – correct or incorrect. Good reliability does not imply that the test has a high sensitivity and specificity, although most likely an unreliable test neither will be sensitive nor specific.¹⁶³ We used test-retest as a way to assess reliability across time, both in study III and in study IV.

Self-reported information

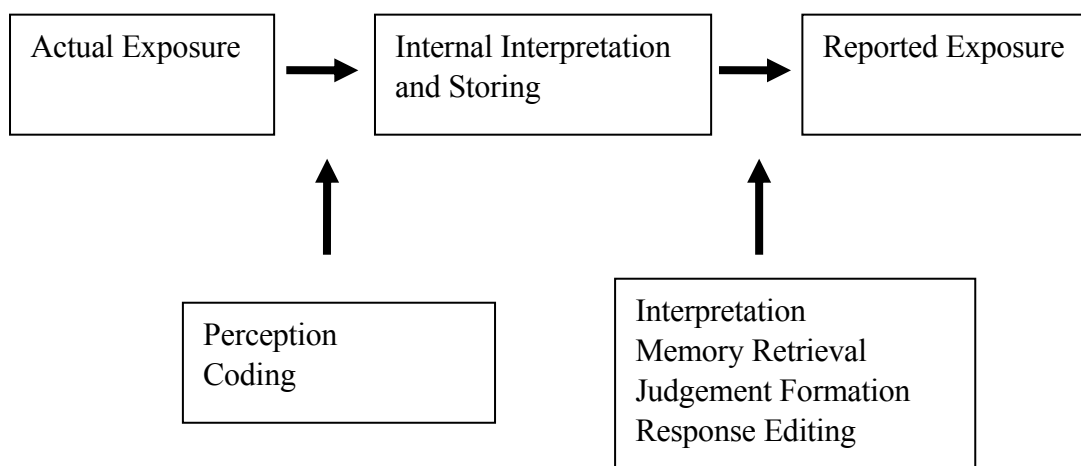


Figure 10. From actual exposure to self-reported exposure.

When we conduct studies based on self-reports we make two assumptions. First we assume that the individual’s internal interpretation of the actual exposure is correct, secondly we assume that the self-reported exposure corresponds to the individual’s correct view (personal communication, Professor Anders Kjellberg). Several threats to validity lurk in the process between actual and reported exposure.

The first step between the actual exposure and the internal interpretation / storing may be threatened by factors we can do little about. See figure 10. To successfully integrate the entire experience of the exposure, in our case, physical activity, and to make an internal interpretation of it and store it, may depend on various factors such as perceptual and cognitive ability or memory capacity. Other factors such as clear physiological effects of a certain exposure (perspiration or breathlessness for example) or the respondent’s personal hypothesis concerning a risk factor and outcome, can skew how well the entire experience of physical activity is perceived, encoded, stored and eventually retrieved when asked for.

In our first validation study, study III, we only asked us if it was at all possible to self-estimate concurrent activity. The step in between, the personal internalization of the workload, was not subject to some of the abovementioned threats, such as memory capacity. We found validity to be high in the process between the actual concurrent and reported exposure.

When the internal interpretation of the activity is converted to self-reported physical activity in for example a questionnaire, the respondent performs four tasks; interpretation, memory retrieval, judgment formation and response editing.⁵⁴ We may be able to minimize possible threats to validity in this process through careful questionnaire design. Specific examples and particular phrasing can minimize the differential interpretation of the question from subject to subject and between subject and researcher. Constructs may be etic – universally understood across all cultural groups, or emic – culture-bound, understood differently or not at all by other cultural groups.¹⁶⁷ Our instrument, first developed for use within Swedish society, is based on examples well known to most people in Sweden (like skiing downhill and shoveling snow) and may need to be adapted for use in other settings. Questions with ambiguous terms and answer alternatives such as “much worse, worse, approximately the same, better, much better” such as the fitness question in study V, leave the field open for misclassification. Misclassification can also occur when physical activity questionnaires do not cover inactivity and low intensity activities. This can render in misclassification due to lack of suitable answer alternatives or in differential loss to follow up by the sedentary population.¹⁶⁸

In our second validation study, study IV, we found an overestimation of self-rated usual activity. With regard to study III we are confident that the interpretation of the different examples is without any larger error. However, once the information of past physical activity is retrieved (memory retrieval) the respondent goes through the judgment formation process. This involves rating, estimation of frequency and evaluation of the relative importance of conflicting information retrieved from memory.⁵⁴ Telescoping¹⁶⁹ could come into play here. Events outside the time period named could to be drawn into the exposure. This artificially inflates the estimates of how often certain behaviors such as strenuous physical activity are reported. Finally, response editing could be another validity problem. Response editing can be in the form of social desirability which results in a tendency to overestimate desirable behaviors, and underestimate undesirable behaviors.^{54,170} In our study, the natural reaction to defend one’s social image when using self-reported measures may have lead to overestimation of exposure.

Confounding

Confounding is when the effect of the exposure under study is mixed together with the effect of another variable. It is one of the most important threats to the validity. For a factor to be a confounder it needs to fulfil three criteria; it has to be a risk factor for the disease under study, it has to be associated with the exposure under study and it can not be in the causal pathway between the exposure and the disease.¹⁶³ We can prevent confounding in the design, but unlike selection and information bias, we can also control for confounding in the analysis phase of the study. In the design of a study we can randomize, restrict and match, and in the analysis we can stratify and conduct regression analyses. Below are examples of how some of these methods were employed in this thesis.

In the second validation study, study IV, we randomized the participants invited to take part to receive either the traditional printed questionnaire or the web questionnaire. Randomization adjusts for all known and unknown confounders.

In study V, the National March Cohort, stratification was done in order to keep a potentially confounding factor constant within each stratum, but it was also a tool to become acquainted with the data. We stratified as an interim step to examine whether background variables affected the place in the distribution when physical activity was measured in different ways.

In study II, we conducted regression analyses; an efficient way of adjusting for several confounders. Primarily we adjusted for age. We then adjusted for covariates which were considered potential confounders in the study of physical activity and total mortality: age at enrollment, years of education, body mass index, alcohol intake, smoking status, mean number of cigarettes, number of years smoking, and country of origin.

Residual confounding

Even if adjustments for confounding have been made, it is difficult to exclude the potential contribution of residual confounding from the findings. Residual confounding arises when a confounding factor is not measured with sufficient precision. In study II, the Women's Lifestyle and Health Cohort, the possibility of physical activity measurement error and confounding by diet (such as high fruit and vegetable intake, low fat intake and low lipid levels) may have resulted in residual confounding. But, given the robustness of our study findings even after extensive control of covariates, it is unlikely that our results would have been qualitatively different had we controlled for these factors.

Precision

Random error could be due to unexplainable random variation in the data – the influence of chance. Random error leads to lack of precision. P-values and confidence intervals convey information about both precision and size of our estimate.¹⁶³ The width of the confidence interval depends on the level of confidence (as confidence increases, the width of the confidence interval increases), size of sample (as the sample size increases, the width decreases) and variability in the population (as variability increases, the width increases). Thus, precision can be improved by increasing sample size – as in the case with study I, the association of interest became significant when more studies were added. Precision can also be improved by a more efficient study design and improved methods, resulting in more and better information per subject.

If the confidence interval does not include the null value, the results will most likely not be due to chance. The confidence level was chosen to be 95 % in our studies. Thus, we are 95 % confident that the limits of the interval cover the truth. Five times out of a hundred, the interval will not cover the truth. We can not rule out that a positive finding is not due to chance. Nor can we rule out that a null result is due to chance. We might not have had statistical power or sufficient methods to detect a true association. (The power is the probability that we will reject H_0 when the null hypothesis is false.) But factors such as strength of the association, consistency with earlier studies, sub-analyses, biological hypotheses and dose-response relationships should be kept in mind when evaluating the role of chance.

INTERPRETATIONS AND IMPLICATIONS

While there is general consensus that an active and fit way of life have beneficial effects on cardiovascular diseases,^{162,171} our studies have shown that physical activity also substantially reduces breast cancer risk and overall mortality in women. This is in line with the literature suggesting that a sedentary life style is associated with several chronic diseases and conditions.¹⁷²⁻¹⁷⁴

Physical activity might be one of the most important modifiable factors that determine risk of chronic morbidity and mortality and it may also be an important adjuvant to therapies against various diseases in routine healthcare. However, in our efforts to investigate such effects, the shortage of practicable, valid, reliable and sensitive instruments for self-recording of all physical activity and inactivity has been a limiting factor. The lack of sophistication in the measurement of physical activity in epidemiological studies was obvious in the search for studies to include in study I, the meta-analysis of physical activity and breast cancer. Studies could for example inquire solely about frequency of exertion^{20,175} which make it difficult to clarify a possible dose-response relationship, since duration and intensity are unclear.

We suspected that the different precisions in the assessments of physical activity may be a source of heterogeneity contributing to inconsistent results among studies, and to the observed heterogeneity across different categories of a number of covariates in our analysis in study I.

In study II we could conclude that encouragement of physical activity in youth is important, not only because it reduces the risk of future breast cancer as seen in study I, but also because it predicts physical activity patterns in adult life. Second, it is never too late in life to commence a physical activity program, as shown by the strong reduction in mortality risk for physically active women that were inactive in the past. Third, physical activity ought to be maintained throughout life. Mortality rates among Norwegian and Swedish women below age 60 could decrease between 9 and 22 per cent if overall physical activity levels would increase in these populations.

Nevertheless, we used a rather rudimentary method to assess physical activity. We had no information of what kind of activities the women practiced, the intensity or the duration. Our only information was the self-rated total level of activity on a scale between 1 and 5. We could not translate the latter conclusion into an instructive public health message concerning how much physical activity in the population needed to increase in order to decrease the mortality in the population by 9 %. Also, for the world audience, time, place and instrument specific attributable fractions may be of marginal importance. This resulted in the choice to exclude the population attributable fraction (PAF) from the final paper.

Optimally, every study should identify frequency, duration and intensity of physical activity, the product of which we would generate total gross energy expenditure.¹⁷⁶ This would be practical for recommendations and comparisons between studies. Better methods may increase our chances of avoiding misclassification and enhance our

understanding of the association between for physical activity and health. Since physical activity is also a potential confounder, better methods may enhance our ability to control for confounding.

We set out to develop a generic instrument for the assessment of physical activity in epidemiological studies. The nine steps in the instrument were anchored in well known examples of activities chosen from reference lists of absolute intensity. The MET-values given in the lists do not estimate the energy cost accounting for efficiency of movement, body mass, age, sex and environmental conditions. However, the respondent chooses the level relative to their own experience of how the intensity of a performed activity is compared to their own experience of the different examples given. The instrument has been shown to be both reliable and valid.

Validity was assessed both in the laboratory against concurrent activity and in a random sample using 24-hour recalls as our gold standard, study III-IV. An objective method like DLW would have been ideal for validation; however it is difficult and expensive to apply on a large scale in an epidemiological setting. Results from earlier studies¹⁴⁵ have shown that three 24-hour recalls provide an assessment of physical activity behavior comparable to other short-term physical activity assessments that utilized activity monitors. Furthermore, this type of validation has been considered satisfactory in nutritional research. Regardless, the focus of our evaluation was people's ability to integrate their perception of the physical activity level over longer periods into one measure of "usual" activity. Although the problem with correlated errors between the two methods should not be underestimated and could certainly lead to exaggerated correlations, the separated self-reports of "usual" activity and short-term retrospective reports covering the preceding 24 hours seem to indicate that "usual" activity reflects the average of three short-term recalls fairly well. Had the self-reports reflected no more than an individual's way to communicate the physical activity level, we would expect little or no intra-individual variation of the 24 hour reports. However, this variation was substantial (data not shown).

In the last study, study V, we compared measures of physical activity obtained with the new instrument with conventional questions commonly used in epidemiological studies. We could conclude that different instruments seem to measure different components of physical activity. Leisure time and household activities represented no more than 17 % of total activity measured by our instrument and correlations were low between different ways to assess physical activity.

Physical activity patterns have traditionally differed between men and women. While women have taken a larger share of household work and child care, men have generally engaged more in heavy leisure time activity.¹ This was reported in our study as well. As conventional physical activity questionnaires have predominantly probed high intensity activities, it is conceivable that misclassification of total physical activity may have affected men and women differentially. However, gender was not significantly linked to leisure time and household physical activity in our multivariable linear regression modeling in study V.

Finally we can conclude that assessment methods of physical activity can be improved. We have provided one model to do so, and it should be seen as proof-of-principle. Our new method seems to be a practical way to collect data, although further methodological refinements are within reach. Clearly open-ended durations work better than predefined time intervals, which tend to result in an underestimation of total duration. Other response alternatives might be even better. A pie chart, in which the participants are asked to display the contribution of each intensity level by dividing the pie accordingly, is one possibility.

9. ONGOING AND FUTURE RESEARCH

In our quest to expand the understanding of how physical activity can protect from breast cancer and premature mortality we realized there was a need for a valid and reliable instrument for self-reported physical activity on an interval scale. The resulting instrument is validated in relation to both concurrent activity and 24-hour recalls. It has been shown to be well suited for use in the general population and in self-administered questionnaires, both printed and electronic.

Our instrument is currently in use in several ongoing studies, among them, the follow-up of the Women's Lifestyle and Health Cohort, which until recently only measured physical activity on a 5-point scale.

When physical activity is measured with better precision, we can improve our understanding of possible mechanisms through which physical activity impacts different biologic systems. Recently researchers at the Department of Epidemiology and Biostatistics, Karolinska Institutet and the Swedish Institute of Infectious Disease Control, conducted a study using the new instrument and immunological markers in blood to examine the relationship between physical activity and immunologic function. Furthermore, a large population study investigating physical activity using the new instrument, and susceptibility to upper respiratory tract infection, as a marker for immune function, is currently underway from the same research group.

With better methods to measure physical activity we have the possibility to conduct randomized controlled trials to investigate the biologic biomarkers associated with various types, intensities and durations of physical activity. These provide ways to better understand the mechanistic pathways through which physical activity may impact the development of cancer and other diseases. Such studies could answer remaining questions about type and amount of activity required to cause a protective effect, whether there are critical time periods when physical activity is more important, or whether physical activity can be modified by other factors. This information is needed to create good quality population-based prevention programmes.¹⁷⁷

Within this thesis we concluded that physical activity in adolescence reduces future breast cancer risk. Physical activity is likely to do so through a hormonally mediated pathway. As a next step, a future exercise intervention clinical trial could examine the effect of different amounts of physical activity in pre and postmenopausal women: the effect on serum sex hormones, metabolic hormones, weight, adipose tissue stores and immune function. We can furthermore investigate if the potential relation between physical activity and endogenous hormones are mediated by abdominal fat. Such a trial, running for at least a year, could include women who are sedentary (<1 hour of moderate physical activity per week at baseline) overweight (BMI >25), in good health, who are not pregnant or taking any hormonal treatment. At baseline we could assess usual physical activity with our instrument and aerobic capacity with a fitness test. We would measure BMI, waist-hip ratio, bioelectric impedance, and intra-abdominal fat using ultrasound or CT. Baseline characteristics would be obtained in a questionnaire and blood and urine samples collected to analyse biomarkers. The participants would

be randomized with equal probability to different doses of leisure time physical activity. This could, for example, be a yoga group, a walking group and a group regularly taking part in supervised aerobic training. Adherence to the prescribed intervention could be measured with class attendance, activity logs and pedometers. At three, six and nine months a shorter questionnaire for example at the same time as our instrument for usual physical activity is administered, biomarkers, and anthropometric measures could be collected. At end of follow-up, after 12 months, a visit with a largely similar data collection as at baseline would be conducted to study the effect of different amounts of physical activity on serum sex hormones, metabolic hormones, weight, adipose tissue stores, immune function and well-being.

Exercise interventions are expensive and time consuming. The suggested study above uses intermediate outcome comparisons for evaluation, instead of comparing the future rate of breast cancer. This is a way to reduce sample size, study duration and cost.¹⁷⁸ But if we need to increase power or decrease cost, we could choose to limit the study to only postmenopausal women. This would render fewer problems associated with the measurements of hormones with timing of the menstrual cycle. Furthermore, the highly metabolic abdominal fat is an important source for estrogens, particularly in postmenopausal women;^{179,180} since the intervention is assumed to result in weight loss, we are more likely to affect the relative concentrations of estrogens in postmenopausal women compared to premenopausal women. Finally, previous investigations have more strongly supported the association between increased physical activity and reduced breast cancer in postmenopausal years.¹⁸¹

The development of the instrument should be seen as proof-of-principle. The examples in the instrument include activities (for example snow shoveling) that may not be typical and well known in all cultures. Examples and intervals could easily be adapted to fit a different setting by constructing an instrument using methods similar to ours. The ability to validly estimate intensity of physical activity should be universal. Our instrument has been adapted for the Danish population.

However, there is still an aspect of validity that warrants further study. A lifetime retrospective history of physical activity is often of interest, particularly in cancer research. But how well is physical activity in the past reported? More needs to be learned about perceiving, encoding, storing and retrieving past levels of physical activity.⁹ Few epidemiological studies have investigated this issue, but there seems to be a bias in accurately recalling activities of different types.^{60,182,183} A study of long and short term memory of physical activity was prepared as a part of this thesis. The study is approved by the ethics committee, conducting it will be the next step.

10. CONCLUSIONS

- Moderate to vigorous physical activity during adolescence/young adulthood reduces breast cancer risk in females. (Study I)
- Total physical activity reduces overall mortality in women – regardless of activity level in the past. (Study II)
- Due to lack of sophistication in the measurement of physical activity in epidemiological studies, misclassification of physical activity is substantial. (Study V)
- Assessment of physical activity in epidemiological studies can be improved. (Study III-V)

11. SAMMANFATTNING (SUMMARY IN SWEDISH)

Bakgrund: Trots oprecisa mätinstrument har ett antal studier visat att fysisk inaktivitet, fastän inte alltid otvetydigt, ökar risken för flera cancerformer och för tidig död. Felklassificering, som följer av dåliga mätmetoder, leder i regel till en underskattning av sambanden. Av den anledningen är den sanna skyddande effekten av fysisk aktivitet sannolikt större och möjligen omfattande flera cancerformer än vad som hittills framkommit. Det finns således en stor potential för cancerprevention i framtiden. Detta förutsätter givetvis att det finns ett orsakssamband som inte är beroende på samvariation med andra riskfaktorer. Fördjupade studier av den fysiska aktivitetens roll är alltså mycket angelägna.

I. Fysisk aktivitet under tonåren minskar risken för bröstcancer – men vår studie visar på stora skillnader i hur man mäter fysisk aktivitet

Trots att ett samband mellan fysisk aktivitet i tonåren och bröstcancer är biologiskt sannolikt, har publicerade studier givit varierande resultat. Vi genomförde en statistisk meta-analys av 23 studiers analysresultat och fann en riskminskning på nästan 20 % (RR 0,81 CI 0,73–0,89) för de tonårsflickor som regelbundet var fysiskt aktiva jämfört med stillasittande flickor. Varje extra timmes motion per vecka minskade risken för bröstcancer senare i livet med 3 %. Använda frågekonstruktioner var genomgående oprecisa och hade ifrågasättbar validitet (ett uttryck för i vilken grad ett test, eller i det här fallet ett formulär, förmår mäta vad det är avsett att mäta). Detta kan vara en av anledningarna till de skiftande resultaten i olika studier.

II. Fysisk aktivitet i arbetsför ålder minskar risken för tidig död för kvinnor – men resultaten är svåröversatta till folkhälsobudskap

För 12 år sedan fyllde 99 099 svenska och norska kvinnor i åldrarna 30–49 år i en enkät om hälsa och livsstilsfaktorer som fysisk aktivitet. Vi har nu följt upp dessa kvinnor via SCB: s befolkningsregister och funnit att risken att dö före 60 års ålder var halverad (RR 0,46 CI 0,33–0,65) bland de mest fysiskt aktiva jämfört med de inaktiva kvinnorna. Vi fann också att skyddseffekten försvann relativt snart om man senare i livet blivit inaktiv. I studien fick kvinnorna skatta sin aktivitet på en skala från 1–5, vilket tyvärr gör det svårare att översätta resultaten till ett tydligt folkhälsobudskap.

III. Utveckling av ett nytt instrument för mätning av fysisk aktivitet i epidemiologiska studier – en studie i laboratoriet

Ospecifika mätningar som gjorts med gamla mätmetoder ger alltså indikationer på att fysisk aktivitet kan vara starkt skyddande – om än kortvarigt – mot såväl bröstcancer som total mortalitet. Emellertid skulle det verkliga genomslaget inom forskningsfältet påskyndas om det fanns bra mätinstrument som med ett begränsat antal frågor fångar upp all aktivitet (även inaktivitet) och där resultaten kan omvandlas till energiåtgång på intervallskalenivå. Vi har utvecklat ett sådant instrument som nu används i en rad studier.

Konstruktionen av vårt instrument har gått till på följande sätt: all fysisk aktivitet delas in i nio intensitetsnivåer, från sömn och vila till hårdast tänkbara ansträngning. En ökning från en nivå till nästa motsvarar en ökning av syreförbrukningen med 3,5 ml O₂

$\times \text{kg}^{-1} \times \text{min}^{-1}$ (1 MET) dvs. skalstegen är multipler av den basala energiförbrukningen i vila. Vid registreringen ska deltagarna ange hur många timmar och minuter de ägnar åt respektive nivå under en typisk dag. De ska alltså slå ihop tiden för olika aktiviteter med samma grad av ansträngning. Intensitetsnivåerna måste vara entydiga och lätta att förstå för alla. Därför definieras de med en rad exempel, som måste vara sådana att de uppfattas på samma sätt av alla människor. För att hitta sådana exempel angav 94 vuxna försökspersoner med skilda bakgrunder sin uppfattning om ett stort antal aktiviteter, vars energiförbrukning tidigare blivit uppmätt. Vi använde visuella analogskalor och med hjälp av vissa förutbestämda kriterier kom det slutliga instrumentet att för varje intensitetsnivå ha fyra exemplifierande aktiviteter.

I nästa steg ville vi fastställa om man kan kvantifiera fysisk aktivitet så att rapporterna verkligen speglar energiförbrukningen. I studien deltog 80 försökspersoner i åldern 16–65, vilka slumpades till fyra olika belastningar på en testcykel, varefter de skattade belastningen med hjälp av instrumentet. Korrelationen (ett statistiskt mått på två variablers samvariation) var god, i genomsnitt 0,89 av 1,0 möjliga. Vid test-retest, ett sätt att mäta reliabiliteten (upprepbarheten), var korrelationskoefficienten 0,99. Således är det möjligt att tillförlitligt självskatta sin fysiska aktivitet med det nya instrumentet.

IV. Utveckling av ett nytt instrument för mätning av fysisk aktivitet i epidemiologiska studier – en studie bland slumpvis utvalda personer ur befolkningen

Därefter studerade vi om man kan sammanfatta en varierande fysisk aktivitet över många dagar och minnas den på ett korrekt sätt. Med hjälp av det nya instrumentet fick 765 slumpmässigt utvalda män och kvinnor i åldrarna 20–59 uppge hur fysiskt aktiva de är under ett vanligt dygn. Därefter blev halva gruppen vid tre tillfällen intervjuad per telefon om hur fysiskt aktiva de varit det senaste dygnet, andra halvan fyllde i instrumentet igen vid tre separata tillfällen. Korrelationskoefficienten mellan total fysisk aktivitet mätt med instrumentet jämfört med genomsnittet för de tre dygnen var 0,73. Reproducerbarheten visade en korrelationskoefficient på 0,55. Det nya instrumentet kan därmed anses mäta total fysisk aktivitet relativt väl.

V. Användningen av ett nytt instrument för mätning av fysisk aktivitet i epidemiologiska studier – resultat från en kohortstudie

Slutligen använde vi instrumentet i en studie som genomfördes i samband med en av Cancerfondens Riksmarscher. Totalt fyllde 43 880 män och kvinnor i det 32-sidiga frågeformuläret som på ett detaljerat sätt mätte olika aspekter på fysisk aktivitet och andra levnadsvanor. Deltagarna hade under ett genomsnittsdygnet en energiförbrukning på 2 353 kcal. Den fysiska aktiviteten sjönk med stigande ålder, BMI och utbildning. Motionsrelaterad aktivitet stod för 17 % av den totala fysiska aktiviteten. Korrelationen mellan motionsrelaterad aktivitet mätt med vanligt förekommande enkätfrågor och all aktivitet mätt med instrumentet var låg (0,26). Korrelationen mellan självskattad kondition respektive självskattad total aktivitet jämfört med andra av samma kön och ålder (ett annat vanligt sätt att mäta fysisk aktivitet i enkäter) och all aktivitet mätt med instrumentet var ännu lägre (korrelationskoefficienterna uppmättes till 0,19 respektive 0,20). Epidemiologiska metoder för att mäta fysisk aktivitet mäter således helt olika saker.

Betydelse: Korrekt information om eventuella skyddseffekter är en förutsättning för att kraftfullt kunna marknadsföra fysisk aktivitet som profylax mot ohälsa och för tidig död. Det nyutvecklade själv-rapporteringsinstrumentet, som mäter total fysisk aktivitet, har en potential för att berika den fortsatta forskningen. Fysisk aktivitet skulle kunna bli nästa stora preventionsinstrument.

12. ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those friends and colleagues who inspired, encouraged, and helped me before and during the process of writing this thesis. In particular I would like to thank:

Olof Nyrén, my advisor, for the privilege of being your PhD student, despite my area of interest. For inspiring discussions, which could get me going for weeks, for sharing your scientific brilliance and wonderful feeling for writing, for your sense of perfection and for the fun we have had on the way!

Hans-Olov Adami, the founder of the Department of Medical Epidemiology and Biostatistics (MEB), for the exciting and unforgettable months we worked together with the Riksmarsch project, for knowing how to trigger me, for encouraging me to apply for Harvard School of Public Health, and for supporting me all the times I went back to Harvard to speed up the process of different projects.

Lorelei Mucci, looking back on my time as a doctoral student, the intense days we worked together on studies II and IV were the most fun, inspiring and productive of them all! You are a great role model and my epidemiological heroine!

Cung-Cheng Hsieh, working with you, the professor whose lectures I enjoyed the most at Harvard, was a dream come true – twice! Your extensive knowledge, your humble spirit and your generosity make you a wonderful role model in the world of research.

Dimitrios Trichopoulos, thank you for sharing your deep knowledge and vast experience in epidemiology, making the process of my second paper an unforgettable experience, but also for friendly comments and kind encouragement on the way. When I listen to you, I can only agree with Goethe who 200 years ago said, “In the limitations he first shows himself the master”.

Stephan Rössner, who 20 years ago inspired me, a small town overweight teenage girl, to set an MD and a PhD as my goals. Letting me be a part of your staff and your shadow wherever you went during some of the most exciting months of my life was invaluable for my dedication and for making it into medical school.

Erik Gylfe, for your devotion to start the Research Oriented Medical School at Uppsala University, and for giving me the opportunity to take part in this programme.

Göran Elinder, who believed in the importance of bringing research into the clinic, and thus gave me the chance to do my residency at Södersjukhuset, Stockholm, alternating financed research with clinical work.

Katarina Bälter, for good friendship and valuable discussions about science, literature, gender issues, children and future and for generously letting me do a validation study within your web study.

Elisabete Wiederpass, my globe trotting coauthor on the second paper, faster than anyone to answer emailed questions. Lisa taught me “the perfect is enemy of the published” – which I should have learned much earlier!

Rino Bellocco, for making research a thriller, but always coming up with the solution on the last page!

Elinor Fondell, who successfully fulfilled the complicated task of coordinating study IV – launched the very same week I gave birth to twins. It was a relief to know the study was in such good hands!

Gunilla Sonnebring, for generous help in many ways – from cats to texts – but most of all for being a wonderful and stable friend at MEB.

Kristina Glimsjö (Eriksson) for your excellent and skilful support with the summer school and later on with the mortality manuscript. I felt very fortunate the times I worked with you.

Kerstin Linderholm, it has been wonderful to discuss boring administrative issues with you. I admire how you in an uncomplicated way solve seemingly hopeless problems with a laugh!

Lasse Bryngelsson at Lukas Production, who did an incredible job in almost no time laying out the questionnaire for the National March Cohort. Your continuous support making different versions of the instrument has been most valuable in the work of this thesis.

Jacob Forsell, for generously sharing your great photos of the National March Cohort!

Stephanie Bakaysa, for your late night proof reading and your feeling for editing. Thanks!

Mukachilima Chikuba from Zambia, **Annamaria Kiure**, from Tanzania and **Pamela Surkan** from the US, my former study group at Harvard School of Public Health, for the hundreds of hours we spent together trying to solve the mystery of epidemiology. You made my year at Harvard unforgettable! I have missed our discussions and your clever input during the work with my thesis. Special thanks to Pam, for sharing both my American and Swedish life, for sharing the fun and the difficulties of thesis work and for being a colleague at MEB!

Anna Bergström, for solving my weird computer problems in the beginning of my thesis work, for our discussions, lunches, giggles, and for time we spent together during our first maternity leaves.

Katja Fall, **Estelle Naumburg**, and **Karin Ekström** friends and colleagues at MEB of whom I especially would like to acknowledge for all the friendship, support and fun

during the years. But also thanks to **Johanna A, Anna C, Anna-Mia E, Elisabeth E, Michael F, Lena G, Catarina J, Mårten K, Cecilia M, Tahereh M, Anna N, Michaela P, Lena R, Sara W** and **Karin W**. I have missed you at MEB.

Cat Halthur, my baby daughters' best friend at MEB, whose wonderful baby entertainment skills gave me valuable minutes to work during my maternity leave.

Åse Troedson, my best friend since half my life, for sharing all the ups and downs with this thesis and in the rest of my life. What would I do without you?

Ingela Trolle, my dear sister, who believed in each and every one of the manuscripts by buying a vintage wine in the start of the process, leaving it to mature till a day in the future when the manuscript would be accepted!

Henrik Trolle, my fun brother who is the witty illustrator of my various course materials.

Sten Trolle, my father and my sounding board, who, regardless of any other commitments, is always available for emergency discussions about anything, and my mother **Gunnela** for love and caring.

Vincent and **Baltzar**, my sons, and **Wendela**, my daughter, for the joy you give me and for all the baby company I have had in front of the computer at MEB and at home.

Johan Lagerros, my dear husband (who decided to follow me to Harvard after only having known me for a week!) for our valuable statistical discussions, your emergency computer support, and scientific advice at any time of the day or night, ever since I met you. Last but not least, your generous love and understanding make my life wonderful!

13. REFERENCES

1. CDC. Physical Activity and Health: A Report of the Surgeon General. Atlanta, GA: USA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.
2. CDC. An Introduction to Epidemiology. Excite. Vol. 2006. Atlanta, USA: Centers for Disease Control and Prevention, 2005.
3. Taylor HL, Klepetar E, Keys A, Parlin W, Blackburn H, Puchner T. Death rates among physically active and sedentary employees of the railroad industry. *Am J Public Health* 1962;52:1697-707.
4. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985;100(2):126-31.
5. Bassett DR, Jr. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport* 2000;71(2 Suppl):S30-6.
6. IARC. IARC Handbooks of Cancer Prevention: Weight Control and Physical Activity. Vol. 6. Lyon: IARC Press, 2002.
7. Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Med Sci Sports Exerc* 2001;33(6 Suppl):S364-9; discussion S419-20.
8. Guyton AC. Textbook of medical physiology. eight edition ed. United States of America: W.B. Saunders Company, Harcourt Brace Jovanovich, Inc., 1991.
9. Lamonte MJ, Ainsworth BE. Quantifying energy expenditure and physical activity in the context of dose response. *Med Sci Sports Exerc* 2001;33(6 Suppl):S370-8; discussion S419-20.
10. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Jr., Montoye HJ, Sallis JF, Paffenbarger RS, Jr. Compendium of physical activities: classification of energy costs of human physical activities [see comments]. *Med Sci Sports Exerc* 1993;25(1):71-80.
11. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, O'Brien WL, Bassett DR, Jr., Schmitz KH, Emplainscourt PO, Jacobs DR, Jr., Leon AS. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;32(9 Suppl):S498-504.
12. Shephard RJ. Absolute versus relative intensity of physical activity in a dose-response context. *Med Sci Sports Exerc* 2001;33(6 Suppl):S400-18; discussion S419-20.
13. McTiernan A, Kooperberg C, White E, Wilcox S, Coates R, Adams-Campbell LL, Woods N, Ockene J. Recreational physical activity and the risk of breast cancer in postmenopausal women: the Women's Health Initiative Cohort Study. *Jama* 2003;290(10):1331-6.
14. 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(9):989-99.

15. Liu S, Lee IM, Linson P, Ajani U, Buring JE, Hennekens CH. A prospective study of physical activity and risk of prostate cancer in US physicians. *Int J Epidemiol* 2000;29(1):29-35.
16. Lamb KL, Brodie DA. Leisure-time physical activity as an estimate of physical fitness: a validation study. *J Clin Epidemiol* 1991;44(1):41-52.
17. Manson JE, Hu FB, Rich-Edwards JW, Colditz GA, Stampfer MJ, Willett WC, Speizer FE, Hennekens CH. A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. *N Engl J Med* 1999;341(9):650-8.
18. 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(2):191-201.
19. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J Clin Epidemiol* 1988;41(2):105-14.
20. Luoto R, Latikka P, Pukkala E, Hakulinen T, Vihko V. The effect of physical activity on breast cancer risk: a cohort study of 30,548 women. *Eur J Epidemiol* 2000;16(10):973-80.
21. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med* 2003;37(3):197-206; discussion 206.
22. American College of Sports Medicine position stand. The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness in healthy adults. *Med Sci Sports Exerc* 1990;22(2):265-74.
23. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, Buchner D, Ettinger W, Heath GW, King AC, et al. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine [see comments]. *Jama* 1995;273(5):402-7.
24. Kesaniemi YK, Danforth E, Jr., Jensen MD, Kopelman PG, Lefebvre P, Reeder BA. Dose-response issues concerning physical activity and health: an evidence-based symposium. *Med Sci Sports Exerc* 2001;33(6 Suppl):S351-8.
25. Pivarnik JM, Reeves MJ, Rafferty AP. Seasonal variation in adult leisure-time physical activity. *Med Sci Sports Exerc* 2003;35(6):1004-8.
26. Matthews CE, Freedson PS, Hebert JR, Stanek EJ, 3rd, Merriam PA, Rosal MC, Ebbeling CB, Ockene IS. Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. *Am J Epidemiol* 2001;153(2):172-83.
27. Levin S, Jacobs DR, Jr., Ainsworth BE, Richardson MT, Leon AS. Intra-individual variation and estimates of usual physical activity. *Ann Epidemiol* 1999;9(8):481-8.
28. Merrill RM, Shields EC, White GL, Jr., Druce D. Climate conditions and physical activity in the United States. *Am J Health Behav* 2005;29(4):371-81.

29. Ma Y, Olendzki BC, Li W, Hafner AR, Chiriboga D, Hebert JR, Campbell M, Sarnie M, Ockene IS. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. *Eur J Clin Nutr* 2005(advance online publication).
30. Ainsworth BE, Montoye HJ, Leon AS. Methods of assessing physical activity during leisure and work. In: Bouchard C, Shepard RJ, Stephens T, eds., eds. *Physical Activity, fitness and health: international proceedings and consensus statement*. Champaign, IL: Human Kinetics, 1994;146-59.
31. Franks PW, Ravussin E, Hanson RL, Harper IT, Allison DB, Knowler WC, Tataranni PA, Salbe AD. Habitual physical activity in children: the role of genes and the environment. *Am J Clin Nutr* 2005;82(4):901-8.
32. Montgomery C, Reilly JJ, Jackson DM, Kelly LA, Slater C, Paton JY, Grant S. Validation of energy intake by 24-hour multiple pass recall: comparison with total energy expenditure in children aged 5-7 years. *Br J Nutr* 2005;93(5):671-6.
33. Butte NF, Wong WW, Treuth MS, Ellis KJ, O'Brian Smith E. Energy requirements during pregnancy based on total energy expenditure and energy deposition. *Am J Clin Nutr* 2004;79(6):1078-87.
34. 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: the Ely Young Cohort Feasibility Study 1994-5. *Br J Nutr* 1997;78(6):889-900.
35. Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, Day NE. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr* 2003;6(4):407-13.
36. Healey J. Future possibilities in electronic monitoring of physical activity. *Res Q Exerc Sport* 2000;71(2 Suppl):S137-45.
37. LaMonte MJ, Blair SN, Church TS. Physical activity and diabetes prevention. *J Appl Physiol* 2005;99(3):1205-13.
38. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc* 2001;33(6 Suppl):S379-99; discussion S419-20.
39. Franks PW, Wong MY, Luan J, Mitchell J, Hennings S, Wareham NJ. Non-esterified fatty acid levels and physical inactivity: the relative importance of low habitual energy expenditure and cardio-respiratory fitness. *Br J Nutr* 2002;88(3):307-13.
40. Brage S, Wedderkopp N, Ekelund U, Franks PW, Wareham NJ, Andersen LB, Froberg K. Features of the metabolic syndrome are associated with objectively measured physical activity and fitness in Danish children: the European Youth Heart Study (EYHS). *Diabetes Care* 2004;27(9):2141-8.
41. Franks PW, Ekelund U, Brage S, Wong MY, Wareham NJ. Does the association of habitual physical activity with the metabolic syndrome differ by level of cardiorespiratory fitness? *Diabetes Care* 2004;27(5):1187-93.

42. Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Perusse L, Leon AS, Rao DC. Familial aggregation of VO₂max response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol* 1999;87(3):1003-8.
43. Le Masurier GC, Lee SM, Tudor-Locke C. Motion sensor accuracy under controlled and free-living conditions. *Med Sci Sports Exerc* 2004;36(5):905-10.
44. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Med* 2001;31(2):91-100.
45. Plasqui G, Joosen AM, Kester AD, Goris AH, Westerterp KR. Measuring free-living energy expenditure and physical activity with triaxial accelerometry. *Obes Res* 2005;13(8):1363-9.
46. Bouten CV, Westerterp KR, Verduin M, Janssen JD. Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Med Sci Sports Exerc* 1994;26(12):1516-23.
47. Ekelund U. Assessment of physical activity and energy expenditure in adolescents. Karolinska Institutet, 2002.
48. Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ. Reliability and validity of the combined heart rate and movement sensor Actiheart. *Eur J Clin Nutr* 2005;59(4):561-70.
49. Zhang K, Pi-Sunyer FX, Boozer CN. Improving energy expenditure estimation for physical activity. *Med Sci Sports Exerc* 2004;36(5):883-9.
50. Pereira MA, FitzerGerald SJ, Gregg EW, Joswiak ML, Ryan WJ, Suminski RR, Utter AC, Zmuda JM. A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc* 1997;29(6 Suppl):S1-205.
51. Borg GA. Perceived exertion. *Exerc Sport Sci Rev* 1974;2:131-53.
52. Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. *Scan J Work Environ Health* 1990;16:55-8.
53. Lee IM, Sesso HD, Oguma Y, Paffenbarger RS, Jr. Relative intensity of physical activity and risk of coronary heart disease. *Circulation* 2003;107(8):1110-6.
54. Warnecke RB, Johnson TP, Chavez N, Sudman S, O'Rourke DP, Lacey L, Horm J. Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol* 1997;7(5):334-42.
55. Marcus PM, Newman B, Moorman PG, Millikan RC, Baird DD, Qaqish B, Sternfeld B. Physical activity at age 12 and adult breast cancer risk (United States). *Cancer Causes Control* 1999;10(4):293-302.
56. Verloop J, Rookus MA, van der Kooy K, van Leeuwen FE. Physical activity and breast cancer risk in women aged 20-54 years. *J Natl Cancer Inst* 2000;92(2):128-35.
57. Villeneuve PJ, Morrison HI, Craig CL, Schaubel DE. Physical activity, physical fitness, and risk of dying. *Epidemiology* 1998;9(6):626-31.
58. Smith GD, Shipley MJ, Batty GD, Morris JN, Marmot M. Physical activity and cause-specific mortality in the Whitehall study. *Public Health* 2000;114(5):308-15.

59. Ainsworth BE, Jacobs DR, Jr., Leon AS. Validity and reliability of self-reported physical activity status: the Lipid Research Clinics questionnaire. *Med Sci Sports Exerc* 1993;25(1):92-8.
60. Jacobs DR, Jr., Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc* 1993;25(1):81-91.
61. Albanes D, Conway JM, Taylor PR, Moe PW, Judd J. Validation and comparison of eight physical activity questionnaires. *Epidemiology* 1990;1(1):65-71.
62. Vaz de Almeida MD, Graca P, Afonso C, D'Amicis A, Lappalainen R, Damkjaer S. Physical activity levels and body weight in a nationally representative sample in the European Union. *Public Health Nutr* 1999;2(1A):105-13.
63. Levi F, Pasche C, Lucchini F, La Vecchia C. Occupational and leisure time physical activity and the risk of breast cancer. *Eur J Cancer* 1999;35(5):775-8.
64. Taylor HL, Jacobs DR, Jr., Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis* 1978;31(12):741-55.
65. Jacobs DR, Jr., Hahn LP, Haskell WL, Pirie P, Sidney S. Reliability and validity of a short physical activity history: CARDIA and the Minnesota Heart Health Program. *J Cardiopul Rehabil* 1989;9:448-459.
66. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr* 1982;36(5):936-42.
67. Ramazzini B. Om arbetares sjukdomar. Åkersberga, Sweden: Arbetsmiljöförlaget, 1991 (1700).
68. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol* 2001;2(3):133-40.
69. Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? *J Cell Mol Med* 2005;9(1):208-21.
70. Chen EC, Brzyski RG. Exercise and reproductive dysfunction. *Fertil Steril* 1999;71(1):1-6.
71. Swerdlow AJ, Schoemaker MJ, Higgins CD, Wright AF, Jacobs PA. Cancer incidence and mortality in men with Klinefelter syndrome: a cohort study. *J Natl Cancer Inst* 2005;97(16):1204-10.
72. Friedenreich CM, Thune I. A review of physical activity and prostate cancer risk. *Cancer Causes Control* 2001;12(5):461-75.
73. Adami H-O, Hunter D, Trichopoulos D, eds. Textbook of cancer epidemiology. Oxford: Oxford University Press, 2002.
74. Oettle GJ. Effect of moderate exercise on bowel habit. *Gut* 1991;32(8):941-4.
75. Oliveria SA, Christos PJ. The epidemiology of physical activity and cancer. *Ann N Y Acad Sci* 1997;833:79-90.

76. Kayaleh RA, Meshkinpour H, Avinashi A, Tamadon A. Effect of exercise on mouth-to-cecum transit in trained athletes: a case against the role of runners' abdominal bouncing. *J Sports Med Phys Fitness* 1996;36(4):271-4.
77. Robertson G, Meshkinpour H, Vandenberg K, James N, Cohen A, Wilson A. Effects of exercise on total and segmental colon transit. *J Clin Gastroenterol* 1993;16(4):300-3.
78. Coenen C, Wegener M, Wedmann B, Schmidt G, Hoffmann S. Does physical exercise influence bowel transit time in healthy young men? *Am J Gastroenterol* 1992;87(3):292-5.
79. Bingham SA, Cummings JH. Effect of exercise and physical fitness on large intestinal function. *Gastroenterology* 1989;97(6):1389-99.
80. Dhabhar FS. Stress, leukocyte trafficking, and the augmentation of skin immune function. *Ann N Y Acad Sci* 2003;992:205-17.
81. Dhabhar FS, Viswanathan K. Short-term stress experienced at time of immunization induces a long-lasting increase in immunologic memory. *Am J Physiol Regul Integr Comp Physiol* 2005;289(3):R738-44.
82. Nieman DC. Is infection risk linked to exercise workload? *Med Sci Sports Exerc* 2000;32(7 Suppl):S406-11.
83. Matthews CE, Ockene IS, Freedson PS, Rosal MC, Merriam PA, Hebert JR. Moderate to vigorous physical activity and risk of upper-respiratory tract infection. *Med Sci Sports Exerc* 2002;34(8):1242-8.
84. Heath GW, Ford ES, Craven TE, Macera CA, Jackson KL, Pate RR. Exercise and the incidence of upper respiratory tract infections. *Med Sci Sports Exerc* 1991;23(2):152-7.
85. Canzian F, McKay JD, Cleveland RJ, Dossus L, Biessy C, Rinaldi S, Landi S, Boillot C, Monnier S, Chajes V, Clavel-Chapelon F, Tehard B, Chang-Claude J, Linseisen J, Lahmann PH, Pischon T, Trichopoulos D, Trichopoulou A, Zilis D, Palli D, Tumino R, Vineis P, Berrino F, Bueno-de-Mesquita HB, van Gils CH, Peeters PH, Pera G, Ardanaz E, Chirlaque MD, Quiros JR, Larranaga N, Martinez-Garcia C, Allen NE, Key TJ, Bingham SA, Khaw KT, Slimani N, Norat T, Riboli E, Kaaks R. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. *Br J Cancer* 2006;94(2):299-307.
86. Yu H, Jin F, Shu XO, Li BD, Dai Q, Cheng JR, Berkel HJ, Zheng W. Insulin-like growth factors and breast cancer risk in Chinese women. *Cancer Epidemiol Biomarkers Prev* 2002;11(8):705-12.
87. Muti P, Quattrin T, Grant BJ, Krogh V, Micheli A, Schunemann HJ, Ram M, Freudenheim JL, Sieri S, Trevisan M, Berrino F. Fasting glucose is a risk factor for breast cancer: a prospective study. *Cancer Epidemiol Biomarkers Prev* 2002;11(11):1361-8.
88. Giovannucci E, Pollak MN, Platz EA, Willett WC, Stampfer MJ, Majeed N, Colditz GA, Speizer FE, Hankinson SE. A prospective study of plasma insulin-like growth factor-1 and binding protein-3 and risk of colorectal neoplasia in women. *Cancer Epidemiol Biomarkers Prev* 2000;9(4):345-9.

89. Stattin P, Rinaldi S, Biessy C, Stenman UH, Hallmans G, Kaaks R. High levels of circulating insulin-like growth factor-I increase prostate cancer risk: a prospective study in a population-based nonscreened cohort. *J Clin Oncol* 2004;22(15):3104-12.
90. DeLellis K, Rinaldi S, Kaaks RJ, Kolonel LN, Henderson B, Le Marchand L. Dietary and lifestyle correlates of plasma insulin-like growth factor-I (IGF-I) and IGF binding protein-3 (IGFBP-3): the multiethnic cohort. *Cancer Epidemiol Biomarkers Prev* 2004;13(9):1444-51.
91. Chan JM, Stampfer MJ, Giovannucci E, Gann PH, Ma J, Wilkinson P, Hennekens CH, Pollak M. Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study. *Science* 1998;279(5350):563-6.
92. Chokkalingam AP, Pollak M, Fillmore CM, Gao YT, Stanczyk FZ, Deng J, Sesterhenn IA, Mostofi FK, Fears TR, Madigan MP, Ziegler RG, Fraumeni JF, Jr., Hsing AW. Insulin-like growth factors and prostate cancer: a population-based case-control study in China. *Cancer Epidemiol Biomarkers Prev* 2001;10(5):421-7.
93. Yu H, Rohan T. Role of the insulin-like growth factor family in cancer development and progression. *J Natl Cancer Inst* 2000;92(18):1472-89.
94. Irwin ML, McTiernan A, Bernstein L, Gilliland FD, Baumgartner R, Baumgartner K, Ballard-Barbash R. Relationship of obesity and physical activity with C-peptide, leptin, and insulin-like growth factors in breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* 2005;14(12):2881-8.
95. Morimoto LM, Newcomb PA, White E, Bigler J, Potter JD. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: personal and lifestyle factors (United States). *Cancer Causes Control* 2005;16(8):917-27.
96. Kaaks R, Lukanova A. Energy balance and cancer: the role of insulin and insulin-like growth factor-I. *Proc Nutr Soc* 2001;60(1):91-106.
97. Physical activity across the cancer continuum: report of a workshop: review of existing knowledge and innovative designs for future research. *Cancer* 2002;95(5):1134-43.
98. Palazzetti S, Richard MJ, Favier A, Margaritis I. Overloaded training increases exercise-induced oxidative stress and damage. *Can J Appl Physiol* 2003;28(4):588-604.
99. Chang CK, Tseng HF, Hsuuw YD, Chan WH, Shieh LC. Higher LDL oxidation at rest and after a rugby game in weekend warriors. *Ann Nutr Metab* 2002;46(3-4):103-7.
100. Finaud J, Lac G, Filaire E. Oxidative stress : relationship with exercise and training. *Sports Med* 2006;36(4):327-58.
101. Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. *Int J Cancer* 1999;83(1):18-29.
102. Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA. Global trends in breast cancer incidence and mortality 1973-1997. *Int J Epidemiol* 2005;34(2):405-12.

103. Parkin DM. International variation. *Oncogene* 2004;23(38):6329-40.
104. Cancer Incidence in Sweden 2003. Stockholm: The National Board of Health and Welfare, Centre for Epidemiology, 2004.
105. Boffetta P, Parkin DM. Cancer in developing countries. *CA Cancer J Clin* 1994;44(2):81-90.
106. Parkin DM. Cancer in developing countries. *Cancer Surv* 1994;19-20:519-61.
107. Deapen D, Liu L, Perkins C, Bernstein L, Ross RK. Rapidly rising breast cancer incidence rates among Asian-American women. *Int J Cancer* 2002;99(5):747-50.
108. Stanford JL, Herrinton LJ, Schwartz SM, Weiss NS. Breast cancer incidence in Asian migrants to the United States and their descendants. *Epidemiology* 1995;6(2):181-3.
109. Ziegler RG, Hoover RN, Pike MC, Hildesheim A, Nomura AM, West DW, Wu-Williams AH, Kolonel LN, Horn-Ross PL, Rosenthal JF, Hyer MB. Migration patterns and breast cancer risk in Asian-American women. *J Natl Cancer Inst* 1993;85(22):1819-27.
110. Chao A, Connell CJ, Jacobs EJ, McCullough ML, Patel AV, Calle EE, Cokkinides VE, Thun MJ. Amount, type, and timing of recreational physical activity in relation to colon and rectal cancer in older adults: the Cancer Prevention Study II Nutrition Cohort. *Cancer Epidemiol Biomarkers Prev* 2004;13(12):2187-95.
111. Colbert LH, Hartman TJ, Malila N, Limburg PJ, Pietinen P, Virtamo J, Taylor PR, Albanes D. Physical activity in relation to cancer of the colon and rectum in a cohort of male smokers. *Cancer Epidemiol Biomarkers Prev* 2001;10(3):265-8.
112. Hsing AW, McLaughlin JK, Chow WH, Schuman LM, Co Chien HT, Gridley G, Bjelke E, Wacholder S, Blot WJ. Risk factors for colorectal cancer in a prospective study among U.S. white men. *Int J Cancer* 1998;77(4):549-53.
113. Wei EK, Giovannucci E, Wu K, Rosner B, Fuchs CS, Willett WC, Colditz GA. Comparison of risk factors for colon and rectal cancer. *Int J Cancer* 2004;108(3):433-42.
114. Steindorf K, Jedrychowski W, Schmidt M, Popiela T, Penar A, Galas A, Wahrendorf J. Case-control study of lifetime occupational and recreational physical activity and risks of colon and rectal cancer. *Eur J Cancer Prev* 2005;14(4):363-71.
115. Littman AJ, Voigt LF, Beresford SA, Weiss NS. Recreational physical activity and endometrial cancer risk. *Am J Epidemiol* 2001;154(10):924-33.
116. Matthews CE, Xu WH, Zheng W, Gao YT, Ruan ZX, Cheng JR, Xiang YB, Shu XO. Physical activity and risk of endometrial cancer: a report from the Shanghai endometrial cancer study. *Cancer Epidemiol Biomarkers Prev* 2005;14(4):779-85.
117. Schouten LJ, Goldbohm RA, van den Brandt PA. Anthropometry, physical activity, and endometrial cancer risk: results from the Netherlands Cohort Study. *J Natl Cancer Inst* 2004;96(21):1635-8.

118. Zeegers MP, Dirx MJ, van den Brandt PA. Physical activity and the risk of prostate cancer in the Netherlands cohort study, results after 9.3 years of follow-up. *Cancer Epidemiol Biomarkers Prev* 2005;14(6):1490-5.
119. Friedenreich CM, McGregor SE, Courneya KS, Angyalfi SJ, Elliott FG. Case-control study of lifetime total physical activity and prostate cancer risk. *Am J Epidemiol* 2004;159(8):740-9.
120. Weiderpass E, Margolis KL, Sandin S, Braaten T, Kumle M, Adami HO, Lund E. Prospective study of physical activity in different periods of life and the risk of ovarian cancer. *Int J Cancer* 2006;118(12):3153-60.
121. Patel AV, Rodriguez C, Pavluck AL, Thun MJ, Calle EE. Recreational Physical Activity and Sedentary Behavior in Relation to Ovarian Cancer Risk in a Large Cohort of US Women. *Am J Epidemiol* 2006.
122. Hahn RA, Teutsch SM, Rothenberg RB, Marks JS. Excess deaths from nine chronic diseases in the United States, 1986. *Jama* 1990;264(20):2654-9.
123. Baik I, Ascherio A, Rimm EB, Giovannucci E, Spiegelman D, Stampfer MJ, Willett WC. Adiposity and mortality in men. *Am J Epidemiol* 2000;152(3):264-71.
124. Blair SN, Kohl HW, 3rd, Barlow CE, Paffenbarger RS, Jr., Gibbons LW, Macera CA. Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *Jama* 1995;273(14):1093-8.
125. Richardson CR, Kriska AM, Lantz PM, Hayward RA. Physical activity and mortality across cardiovascular disease risk groups. *Med Sci Sports Exerc* 2004;36(11):1923-9.
126. Bucksch J. Physical activity of moderate intensity in leisure time and the risk of all cause mortality. *Br J Sports Med* 2005;39(9):632-8.
127. Lee IM, Skerrett PJ. Physical activity and all-cause mortality: what is the dose-response relation? *Med Sci Sports Exerc* 2001;33(6 Suppl):S459-71; discussion S493-4.
128. Oguma Y, Sesso HD, Paffenbarger RS, Jr., Lee IM. Physical activity and all cause mortality in women: a review of the evidence. *Br J Sports Med* 2002;36(3):162-72.
129. Andersen LB, Schnohr P, Schroll M, Hein HO. All-cause mortality associated with physical activity during leisure time, work, sports, and cycling to work. *Arch Intern Med* 2000;160(11):1621-8.
130. Kujala UM, Kaprio J, Sarna S, Koskenvuo M. Relationship of leisure-time physical activity and mortality: the Finnish twin cohort. *Jama* 1998;279(6):440-4.
131. Egger M, Smith GD. Meta-Analysis. Potentials and promise. *Bmj* 1997;315(7119):1371-4.
132. Company HM. The American Heritage® Dictionary of the English Language 2004. Fourth Edition Copyright © ed. Vol. 2006 Houghton Mifflin Company, 2000.
133. Cooper H, Hedges LV. The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994.

134. Jenicek M. Meta-analysis in medicine. Where we are and where we want to go. *J Clin Epidemiol* 1989;42(1):35-44.
135. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *Bmj* 1997;315(7121):1533-7.
136. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. Using meta-analysis for research synthesis: pooling data from several studies. *Biostatistics in clinical medicine* McGraw-Hill Health Professions Division, 1993;332-360.
137. Colditz GA, Frazier AL. Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. *Cancer Epidemiol Biomarkers Prev* 1995;4(5):567-71.
138. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama* 2000;283(15):2008-12.
139. Yarnitsky D, Sprecher E, Zaslansky R, Hemli JA. Multiple session experimental pain measurement. *Pain* 1996;67(2-3):327-33.
140. Flint A, Raben A, Blundell JE, Astrup A. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *Int J Obes Relat Metab Disord* 2000;24(1):38-48.
141. Capodaglio EM. Comparison between the CR10 Borg's scale and the VAS (visual analogue scale) during an arm-cranking exercise. *J Occup Rehabil* 2001;11(2):69-74.
142. Joyce CR, Zutshi DW, Hrubes V, Mason RM. Comparison of fixed interval and visual analogue scales for rating chronic pain. *Eur J Clin Pharmacol* 1975;8(6):415-20.
143. Balter KA, Balter O, Fondell E, Lagerros YT. Web-based and mailed questionnaires: a comparison of response rates and compliance. *Epidemiology* 2005;16(4):577-9.
144. Balter O, Balter KA. Demands on web survey tools for epidemiological research. *Eur J Epidemiol* 2005;20(2):137-9.
145. Matthews CE, Freedson PS, Hebert JR, Stanek EJ, 3rd, Merriam PA, Ockene IS. Comparing physical activity assessment methods in the Seasonal Variation of Blood Cholesterol Study. *Med Sci Sports Exerc* 2000;32(5):976-84.
146. Colditz GA, Berkey CS, Mosteller F. Meta-analysis of randomized trials. In: Hennekens CH, ed. *Clinical trials in cardiovascular disease A companion to Braunwald's heart disease*. Philadelphia: W.B. Saunders Company, 1999;43-51.
147. Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995;142(4):383-7.
148. Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979;86(3):598-604.
149. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50(4):1088-101.

150. Pagano M, Gavreau K. Principles of biostatistics. Belmont, CA: Wadsworth Publishing Company, 1992.
151. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health* 1998;88(1):15-9.
152. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985;122(5):904-14.
153. Bosetti C, Tzonou A, Lagiou P, Negri E, Trichopoulos D, Hsieh CC. Fraction of prostate cancer incidence attributed to diet in Athens, Greece. *Eur J Cancer Prev* 2000;9(2):119-23.
154. Rosner B. Fundamentals of Biostatistics. Fourth ed. The United States of America: Wadsworth Publishing Company, 1995.
155. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45(1):255-68.
156. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.
157. Lin LI. A note on the concordance correlation coefficient. *Biometrics* 2000;56(1):324-25.
158. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999;8:135-160.
159. Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, de Ramos M, Hewitt D, Grambsch PV, Kassim N, Little JA. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr* 1979;32(12):2546-9.
160. Council NHaMR. How to review the evidence: systematic identification and review of the scientific literature. Handbook series on preparing clinical practice guidelines. Canberra: Biotext, 1999.
161. Holloszy JO. Mortality rate and longevity of food-restricted exercising male rats: a reevaluation. *J Appl Physiol* 1997;82(2):399-403.
162. Lee CD, Blair SN, Jackson AS. Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *Am J Clin Nutr* 1999;69(3):373-80.
163. Rothman KJ, Greenland S. Modern Epidemiology. Second ed. Philadelphia, PA: Lippincott-Raven Publishers, 1998.
164. Encyclopedia of Biostatistics. 2nd Edition online version ed John Wiley & Sons, Inc., 2005.
165. Lund E, Kumle M, Braaten T, Hjartaker A, Bakken K, Eggen E, Gram TI. External validity in a population-based national prospective study--the Norwegian Women and Cancer Study (NOWAC). *Cancer Causes Control* 2003;14(10):1001-8.
166. Persson L. Level and field of education 1975-2003. Örebro: Statistics Sweden, Education, 2005.

167. Schallock RL, Verdugo MA, Jenaro C, Wang M, Wehmeyer M, Jiancheng X, Lachapelle Y. Cross-cultural study of quality of life indicators. *Am J Ment Retard* 2005;110(4):298-311.
168. Hays LM, Clark DO. Correlates of physical activity in a sample of older adults with type 2 diabetes. *Diabetes Care* 1999;22(5):706-12.
169. Rubin DC. *Autobiographical memory*. New York, NY: Cambridge University Press, 1988.
170. Hebert JR, Ebbeling CB, Matthews CE, Hurley TG, Ma Y, Druker S, Clemow L. Systematic errors in middle-aged women's estimates of energy intake: comparing three self-report measures to total energy expenditure from doubly labeled water. *Ann Epidemiol* 2002;12(8):577-86.
171. Wendel-Vos GC, Schuit AJ, Feskens EJ, Boshuizen HC, Verschuren WM, Saris WH, Kromhout D. Physical activity and stroke. A meta-analysis of observational data. *Int J Epidemiol* 2004;33(4):787-98.
172. Thune I, Njolstad I, Lochen ML, Forde OH. Physical activity improves the metabolic risk profiles in men and women: the Tromso Study. *Arch Intern Med* 1998;158(15):1633-40.
173. Manson JE, Nathan DM, Krolewski AS, Stampfer MJ, Willett WC, Hennekens CH. A prospective study of exercise and incidence of diabetes among US male physicians. *Jama* 1992;268(1):63-7.
174. Thune I, Furberg AS. Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. *Med Sci Sports Exerc* 2001;33(6 Suppl):S530-50; discussion S609-10.
175. Albanes D, Blair A, Taylor PR. Physical activity and risk of cancer in the NHANES I population. *Am J Public Health* 1989;79(6):744-50.
176. Hardman AE. Physical activity and health: current issues and research needs. *Int J Epidemiol* 2001;30(5):1193-7.
177. Rundle A. Molecular epidemiology of physical activity and cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14(1):227-36.
178. Prentice RL, Willett WC, Greenwald P, Alberts D, Bernstein L, Boyd NF, Byers T, Clinton SK, Fraser G, Freedman L, Hunter D, Kipnis V, Kolonel LN, Kristal BS, Kristal A, Lampe JW, McTiernan A, Milner J, Patterson RE, Potter JD, Riboli E, Schatzkin A, Yates A, Yetley E. Nutrition and physical activity and chronic disease prevention: research strategies and recommendations. *J Natl Cancer Inst* 2004;96(17):1276-87.
179. McTiernan A. Physical activity and the prevention of breast cancer. *Medscape Womens Health* 2000;5(5):E1.
180. Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. *J Mammary Gland Biol Neoplasia* 2002;7(1):3-15.
181. Friedenreich CM. Physical activity and breast cancer risk: the effect of menopausal status. *Exerc Sport Sci Rev* 2004;32(4):180-4.
182. Klesges RC, Eck LH, Mellon MW, Fulliton W, Somes GW, Hanson CL. The accuracy of self-reports of physical activity. *Med Sci Sports Exerc* 1990;22(5):690-7.

183. Buchowski MS, Townsend KM, Chen KY, Acra SA, Sun M. Energy expenditure determined by self-reported physical activity is related to body fatness. *Obes Res* 1999;7(1):23-33.