Validity and Repeatability of a Modified Baecke Questionnaire on Physical Activity

MARGREET A POLS,* PETRA H M PEETERS,* H BAS BUENO-DE-MESQUITA,** MARGA C OCKÉ,** COOSJE A WENTINK,*† HAN C G KEMPER† and HUBERTINE J A COLLETTE*


Background. In a pilot study for the European Prospective Investigation into Cancer and Nutrition (EPIC) a modification of the Baecke questionnaire on physical activity was tested for repeatability and relative validity in a population of 134 men and women aged 20–70 years.

Methods. For the assessment of repeatability Pearson's correlation coefficients and percentages of agreement after classification in tertiles were computed between administrations of the questionnaire at baseline, and after 5 and 11 months. Relative validity was determined by comparing the questionnaire to a four times repeated 3-day activity diary.

Results. Repeatability after 5 and 11 months was good, with test-retest correlation coefficients between 0.65 and 0.89 for main sections of the questionnaire. The percentages of agreement, exceeding chance (Cohen's kappa) were 57% and 56% for men (at 5 and 11 months respectively) and 41% and 46% for women. The correlations with the diaries were 0.56 in men and 0.44 in women. Agreement apart from chance between classification in tertiles for both methods was 35% for men and 10% for women.

Conclusions. These data show that repeatability is good and relative validity as compared to an activity diary is moderate but well within the range of values found in other studies. The questionnaire is more valid in men than in women.

In the past few decades, physical activity is getting more and more attention in epidemiological studies. Relationships have been described between inactivity and several diseases. The most well-known are the cardiovascular diseases, but also some kinds of cancer, diabetes and osteoporosis are thought to have a relationship with lack of physical activity.†

Not all methods for assessing physical activity are appropriate for large population studies. The method most frequently used in this type of study is the questionnaire. Questionnaires can differ in reference period asked about, nature and detail of the activities, mode of data collection (interview or self-administered), and method of computing energy expenditure or an activity index. Before administering a questionnaire in a large population, it is important to assess its validity in the population in which it will be used.

The EPIC study (European Prospective Investigation into Cancer and Nutrition) was started in 1989–1990 in seven European countries. It is a large multi-centre prospective study on the relationship between nutrition and the occurrence of cancer. Approximately 400,000 subjects will be included in the study with about equal numbers of men and women.‡

The subjects are asked to complete a detailed dietary questionnaire and a questionnaire on other lifestyle factors like physical activity, smoking, occupation, and reproductive history. In addition, height, weight, and waist/hip circumference are measured and samples of venous blood are collected. Follow-up will be based on incidence data from hospital and cancer registries and on mortality data from death certificates. In 1991 and 1992, methodological pilot studies were carried out in each of the participating countries to evaluate the validity and repeatability of different questionnaires. In the* Dutch validation study (BALANS), which was conducted from October 1991 to October 1992, a questionnaire on physical activity was tested.

* Department of Epidemiology, Faculty of Medicine, University of Utrecht, PO Box 80035, 3508 TA Utrecht, The Netherlands.
** Department of Chronic Disease and Environmental Epidemiology, National Institute of Public Health and Environmental Protection, PO Box 1, 3720 BA Bilthoven, The Netherlands.
† Department of Health Science, Faculty of Human Movement Sciences, Vrije Universiteit, van der Boechorststraat 7–9, 1081 BT Amsterdam, The Netherlands.
This questionnaire was a modification of the physical activity questionnaire originally developed by Baecke.\(^3\) The Baecke questionnaire was chosen because it is short, easy to fill in, and its validity and repeatability have been described in different populations.\(^3\),\(^4\)

**SUBJECTS AND METHODS**

To assess repeatability, the questionnaire was administered three times: in November 1991, April 1992, and October 1992. To determine relative validity, a generally accepted assessment method of supposedly higher quality was chosen as the reference method; a 3-day activity diary based on the method of Bouchard.\(^5\) The diaries were filled in four times: in April, May, June, and July 1992.

**Subjects**

For the BALANS study, 64 men aged 20–60 years and 62 women aged 20–70 years were recruited from a breast cancer screening programme in Utrecht and a monitoring programme on risk factors for chronic diseases in Amsterdam, Doetinchem and Maastricht. These two ongoing projects were chosen because the Dutch EPIC study will be conducted within these projects. The selection was done in such a way as to ensure an equal distribution of subjects over city, sex, and 20-year age group. All subjects signed an informed consent form. Men had a mean age (± SD) of 41.1 (± 11.0) years, and mean Quetelet Index (QI = weight/height\(^2\)) was 25.4 (± 2.8) kg/m\(^2\). For women, mean age was 48.8 (± 14.8) years, and mean QI 24.8 (± 3.5) kg/m\(^2\).

There were six subjects whose activity patterns changed markedly during the study year: five of them stopped working and one restarted formal training which implied a change in activity pattern. Two subjects dropped out of the study in October 1991, before the first questionnaire was administered. These eight subjects were excluded from all analyses. During the course of the study another 11 subjects dropped out. This, together with a varying number of subjects who missed one or more questions, resulted in a different number of subjects every time the questionnaire was administered. Therefore, in the Tables the number of subjects is given for every measurement.

**Questionnaire**

The physical activity questionnaire tested in this study was a modification of the questionnaire developed by Baecke (Appendix).\(^3\) In the modified version the section on leisure time was expanded by three questions.

In the introduction of the questionnaire, it was explained that the questions referred to physical activity during the past year. Subjects were instructed to consider household activities or studying as their work if that was their main daily activity.

Of the 126 subjects who completed the first questionnaire, six did not fill in the work section, four missed question 8, five question 10, and one missed question 13. The second questionnaire was completed by 119 subjects, but seven did not fill in the work section, one missed question 8, and three missed question 10. Of the 116 subjects who filled in the last questionnaire, seven did not fill in the work section. Questions 8 and 10 were both missed by two subjects, and questions 6, 12, and 13 by one subject.

If the answer to a question was missing, the related index was assumed to be the mean score of the remaining answers of that subject. If a subject did not fill in any of the work questions, e.g. because of retirement, it was assumed that he or she would spend the majority of the extra time as leisure time, rather than in playing sports. Therefore, in those cases, the total index was computed as two times the leisure time index plus the sports index.

**Diary**

The reference method for assessment of the validity of the activity questionnaire was a four times repeated 3-day activity diary. The diary was based on the method of Bouchard.\(^5\)

Subjects registered their activities in periods of 5 minutes, assigning to every period a letter indicating the activity (e.g. R for resting, S for standing). If an activity could not be assigned to one of the coded categories (resting, sitting, standing, walking, cycling, driving a car, housekeeping, dressing), a T or an X was entered. The T stood for training or playing sports and the X for other activities. If T or X was used, the subjects were asked to note at the bottom of the page which activity was meant by T or X.

The average daily energy expenditure was computed using the energy cost for each category as estimated from reference values found in literature.\(^6\) Activities described under T or X were classified into four groups according to energy cost. For each group a mean energy expenditure was estimated and used to compute energy expenditure per 24 hours.

The reference energy values per activity that were used to assess total energy expenditure were the same for men and women and were independent of body weight. The 3-day activity diaries were administered
in April, May, June, and July 1992. Of the 126 participants, 115 filled in activity diaries; 103 subjects completed all 12 days, seven subjects 9 days, three subjects 6 days, and two subjects 3 days.

If less than 12 days were registered, energy expenditure was computed as the mean expenditure of the recorded days.

Statistical Methods
Repeatability was assessed in two ways. First, Pearson’s correlation coefficients were calculated between the first and the second, and between the first and the third times the questionnaire was filled in, for each of the four indices (work, sports, leisure time and total index). Second, subjects were classified in tertiles according to the score of the total index of the three questionnaires. Subsequently, the percentage of agreement, Cohen’s kappa (percentage of agreement corrected for chance), and percentage of gross misclassification were calculated.

The same methods were used for the determination of the relative validity of the questionnaire. Pearson’s correlation coefficients were calculated between the total Baecke index of November 1991 and mean energy expenditure per 24 hours derived from the diaries.

Subjects were classified in tertiles according to total energy expenditure as computed from the diaries. The percentage of agreement, Cohen’s kappa, and percentage of gross misclassification were calculated between tertiles of the total Baecke index of November and tertiles of mean 24-hour energy expenditure obtained from the diaries. By estimating validity of the first questionnaire, any influence of a learning effect is excluded.

The questionnaire measures physical activity in the preceding year, and so the correlation between the first questionnaire and the activity diaries may underestimate the validity of the questionnaire. The third questionnaire measures physical activity in the year during which the diaries were collected. Therefore, the correlation coefficient between the third questionnaire and the diaries was also computed. This correlation coefficient might be an overestimation of validity because the subjects were more aware after recording activity several times. Thus, the true correlation may lie somewhere in between. The 95% confidence intervals for correlation coefficients were determined using Geigy scientific tables.  

Analyses were carried out for men and women separately. Both for the questionnaire and the diaries, tertile classifications were made on the basis of the sex-specific distributions.

RESULTS
In Table 1 mean scores for all indices of the three Baecke questionnaires are presented for men and women, as well as mean energy expenditure per 24 hours calculated from the diaries. The main difference between men and women appears to be in the sports category.

Repeatability
The correlation coefficients between the work, sports, leisure time and total indices of the first and second, and of the first and the third questionnaire were good, ranging from 0.65 to 0.89 (Table 2). Generally, repeatability at 5 months was higher than at 11 months although differences were small. Repeatability for men and women was of the same order of magnitude. Scoring of leisure time activities for men and sports for women were least repeatable, but still ≥0.65.

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Men</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>work</td>
<td>2.6 ± 0.6 (58)</td>
<td>2.6 ± 0.7 (56)</td>
<td>2.6 ± 0.7 (54)</td>
<td></td>
<td>2.7 ± 0.5 (62)</td>
<td>2.7 ± 0.5 (56)</td>
<td>2.6 ± 0.5 (55)</td>
<td></td>
</tr>
<tr>
<td>sports</td>
<td>2.7 ± 0.8 (64)</td>
<td>2.8 ± 0.8 (63)</td>
<td>2.8 ± 0.8 (61)</td>
<td></td>
<td>2.1 ± 0.7 (62)</td>
<td>2.2 ± 0.6 (56)</td>
<td>2.1 ± 0.7 (55)</td>
<td></td>
</tr>
<tr>
<td>leisure</td>
<td>2.8 ± 0.5 (64)</td>
<td>2.9 ± 0.5 (63)</td>
<td>2.9 ± 0.5 (61)</td>
<td></td>
<td>2.6 ± 0.5 (62)</td>
<td>2.7 ± 0.5 (56)</td>
<td>2.7 ± 0.5 (55)</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>8.1 ± 1.2 (64)</td>
<td>8.4 ± 1.2 (63)</td>
<td>8.3 ± 1.3 (61)</td>
<td></td>
<td>7.4 ± 1.0 (62)</td>
<td>7.6 ± 1.1 (56)</td>
<td>7.5 ± 1.1 (55)</td>
<td></td>
</tr>
<tr>
<td>diary</td>
<td>12 204 ± 1478 (60)</td>
<td></td>
<td></td>
<td></td>
<td>11 960 ± 1085 (55)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Mean ± SD; number of subjects in brackets.
When subjects were classified into tertiles according to the total indices in November 1991 and April 1992, 71.4% of men and 60.7% of women were classified into the same tertile. Cohen’s kappa was 57.1% for men and 41.0% for women. Gross misclassification (opposite tertiles) occurred in 0% of men and 5.4% of women.

Comparing the questionnaires of November 1991 and October 1992, 70.5% of men and 63.4% of women were classified into the same tertile (Cohen’s kappa, 55.7 and 45.5% respectively). Gross misclassification occurred in 3.3% and 3.6% of men and women.

**Validity**
A scatterplot of the total Baècke scores against mean energy expenditure per 24 hours estimated from the diaries indicates that there is a linear relationship between the two variables (Figure 1), which justifies the use of correlation coefficients.

The Pearson’s correlation coefficient between the total Baècke index and mean energy expenditure per 24 hours was 0.56 for men and 0.44 for women (Table 3). Correction for age changed the correlation coefficients very little, so the uncorrected coefficients are given.

The percentage of agreement between the classification in tertiles according to the total Baècke index in November 1991 and according to total energy expenditure calculated from the diaries was 56.7% for the male group and 40.0% for the female group. Gross misclassification occurred in 8.3 and 10.9% respectively.

The correlation coefficient between the third questionnaire and the diaries was 0.66 for men and 0.42 for women (percentages of agreement 58.3% and 44.4% respectively).

**DISCUSSION**
Of all invited subjects 25% responded positively to the invitation. The most frequently observed reason for non-participation was lack of time. Data on non-responders are limited. For both men and women non-responders did not differ from responders or participants in age distribution (χ² test). The low participation may be explained by the fact that the study design demanded a lot of time and co-operation from the participants. Apart from the physical activity part, subjects also had to complete a questionnaire on diet three times. They had to visit the research centre several times and were visited at home twice for dietary recalls. In addition, 24-hour urine and blood samples were collected four times. If anything, the relatively heavy schedule of activities may have created selection bias towards more health oriented people, possibly leading to overestimation of the quality of the test instrument.

The sample size was based on the recommendation of Willett² to include 100–200 subjects in the population if the expected correlation between a (dietary) questionnaire and ‘truth’ is 0.5–0.7. It was assumed that the same advice would apply to the validation of a physical activity questionnaire.

Although the study population of the BALANS study was not selected from among the participants of the main cohort study, the sampling procedure, city of residence, and age range were similar, and it is not unreasonable to assume that our findings may be extrapolated to members of the cohort study.

The Baècke questionnaire has advantages which make it feasible for large population studies: it is a short questionnaire that is user-friendly. A disadvantage, however, is the subjective five-point scale.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 months</td>
<td>11 months</td>
</tr>
<tr>
<td>work</td>
<td>0.89; 0.82–0.93 *</td>
<td>0.83; 0.72–0.90</td>
</tr>
<tr>
<td></td>
<td>(56)</td>
<td>(54)</td>
</tr>
<tr>
<td>sports</td>
<td>0.88; 0.81–0.93</td>
<td>0.81; 0.70–0.88</td>
</tr>
<tr>
<td></td>
<td>(63)</td>
<td>(61)</td>
</tr>
<tr>
<td>leisure</td>
<td>0.76; 0.63–0.85</td>
<td>0.71; 0.56–0.82</td>
</tr>
<tr>
<td></td>
<td>(63)</td>
<td>(61)</td>
</tr>
<tr>
<td>total</td>
<td>0.85; 0.76–0.91</td>
<td>0.80; 0.69–0.88</td>
</tr>
<tr>
<td></td>
<td>(63)</td>
<td>(61)</td>
</tr>
<tr>
<td>Tertiles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% agreement</td>
<td>71.4%</td>
<td>70.5%</td>
</tr>
<tr>
<td>kappa</td>
<td>57.1%</td>
<td>55.7%</td>
</tr>
<tr>
<td>gross</td>
<td>0.0%</td>
<td>3.3%</td>
</tr>
<tr>
<td>misclassification</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* r: 95% confidence intervals.
  Number of subjects in brackets.
The sports question is the only question that asks for objective information about frequency, duration, and intensity of activities.

People may be inclined to preferentially opt for the middle categories, decreasing between-person variability and leading to an underestimation of correlation coefficients. Computed over all questions, 33% of the answers were in the middle category. It is unclear whether this represents the real distribution or an overestimation.

Questions 8 and 10 were the questions with the most missing values. Obviously, people find it hard to compare their activity to that of other people. Nevertheless, the numbers of missing answers were small, so it is not likely that outcomes will be biased. In a population of 306 men and women aged 20–32 years, Baecke found, in men, average work, sports and leisure time indices of 2.6, 2.8 and 2.8 and, in women, 2.9, 2.4 and 3.1 respectively.\(^3\) In the BALANS study the scores for men were comparable to Baecke’s findings, but for
women the scores were lower. In the BALANS population the women especially were relatively old (all men were <60 whereas 32% of the women were ≥60 years). The level of physical activity has been found to decrease with age, so the lower scores found in this study might be the result of the higher age of the women. Conversely, elderly women may engage in types of physical activity not covered by the questionnaire, e.g. climbing stairs.

The repeatability at 5 months of this modification of the Baecke questionnaire in this population was good, which implies small random error, no major attenuation of risk estimates due to random error, and, consequently, a greater chance of detecting weak associations, assuming that the method is valid.

It is not clear why the leisure time index in men and the sports index in women showed somewhat lower repeatability. In the male group, the leisure time index had the lowest variability, which may have contributed to a lower correlation coefficient. In the female group, variability of the sports index was about the same as that of the other two indices. Although in the present study the test-retest interval was longer, the correlation coefficients were of the same order of magnitude as found by Baecke et al. who reported a 3-months test-retest reliability of 0.88 for the work score, 0.81 for the sport score and 0.74 for the leisure time score in his study population.

Voorrips et al. found a Spearman’s correlation coefficient of 0.89 when an adaptation of the Baecke questionnaire was administered with a 20-days test-retest interval in a population of 29 men and women aged 63–80 years.

Washburn et al. found a 3–7 week repeatability of 0.75 for a newly developed Physical Activity Scale for the Elderly which is based on 7-day recalls. Cauley et al. reported the repeatability of the Pfaffenbarger questionnaire in a population of postmenopausal women: the 4-week repeatability for three different questions varied from 0.76 to 0.97 while 1-year repeatability varied from 0.42 to 0.73.

The repeatability of the classification into tertiles was moderate but compared well with the results reported by Voorrips et al. who found a percentage of agreement of 72% with a test-retest interval of 20 days.

The activity diary was chosen as a validation instrument because it was considered to be a method with errors uncorrelated with the error of the questionnaire.

To increase the representativeness a diary should include weekend as well as weekdays and different seasons of the year. In the BALANS study, the subjects registered all days of the week but for practical reasons it was not possible to have the subjects fill in the diaries in different seasons. This may have had some effect on absolute energy expenditure but probably not on the ranking of subjects.

The correlation between the total Baecke index and energy expenditure was moderate. The lower correlation in women may be explained by a lower variation in energy expenditure computed from the diaries.

Comparing the third questionnaire with the diaries yielded better results than the first questionnaire, especially in men. This difference could be the result of a learning effect, but an explanation might also be that the third questionnaire represents the time period during which the diaries were collected.

Because of differences in methods or study population, correlation coefficients between physical activity questionnaires and validation instruments vary substantially in the literature. Voorrips et al. found a Spearman’s correlation coefficient of 0.78 with a 24-hour activity recall in a population of 31 men and women aged 63–80 years. Albannes et al. validated eight different questionnaires, including the Baecke questionnaire, against measured energy intake and found correlation coefficients between 0.13 and 0.49 (0.38 for the Baecke questionnaire) in a population of 21 men aged 28–55 years. Cauley et al. evaluated five methods for assessing physical activity in postmenopausal women by comparing the methods and also found mostly poor correlation coefficients (range −0.31 to 0.97 with most correlations between −0.20 and 0.30). So, the correlation coefficients found in the BALANS study were well within the range of values reported in other studies.

The total Baecke index may not be a good instrument for all populations to classify subjects according to level of physical activity. It is possible, with the smaller standard deviation (and coefficient of variation) of the diary energy expenditure in women, that a lot of women had a score close to a tertile boundary and ended up in different tertiles for the first and the second measurement without a substantial change in activity. On the other hand, the poor results in women may have been caused by the fact that the diary was not an optimal reference method in this population.

Our findings indicate that, in men, the level of agreement between the Baecke questionnaire and the diary may be adequate to use this questionnaire in epidemiological studies. The low correlation and the very low kappa, possibly caused by the small standard deviation of the computed energy expenditure, may indicate that the Baecke questionnaire should not be used in women. To classify a population of (elderly) women by level of physical activity, a questionnaire
with more emphasis on household tasks, like Voorrips’ questionnaire,\textsuperscript{10} might be a better instrument.

Conclusions drawn from this study should not be extrapolated automatically to other populations. This study was conducted in a group of healthy Dutch adults. It may be that the questionnaire performs differently in populations of other culture, race, or age, because activity patterns can be very different.\textsuperscript{15}

ACKNOWLEDGEMENTS

The authors thank the ‘Stichting Preventicon voor de Vroege Opsporing van Borstkanker in Midden-Nederland’ for its co-operation. The authors also thank H E Goddijn and A M Jansen, and the interviewers for their help in collecting the data. This study was financially supported by the Europe Against Cancer Programme of the Commission of European Communities (WHO/AEP/90/05).

REFERENCES


(Revised version received September 1994)
APPENDIX

Modified Baecke Questionnaire

1. What is your main occupation? ... 1 - 3 - 5
2. At work I sit ... never/seldom/sometimes/often/always 5 - 4 - 3 - 2 - 1
3. At work I stand ... never/seldom/sometimes/often/always 1 - 2 - 3 - 4 - 5
4. At work I walk ... never/seldom/sometimes/often/always 1 - 2 - 3 - 4 - 5
5. At work I lift heavy loads ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
6. After work I am tired ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
7. At work I sweat ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
8. In comparison with others of my own age I think my work is physically ... much lighter/lighter/as heavy/heavier/much heavier 1 - 2 - 3 - 4 - 5
9. Do you play sport? yes/no If yes:
  - which sport do you play most frequently? ... 1 - 3 - 5
  - how many hours a week? <1/1-2/2-3/3-4/4> / 1-2-3-4-5
  - how many months a year: >1/1-3/3-4/4-5 > 1-2-3-4-5
If you play a second sport:
  - which sport is it? ... 1 - 3 - 5
  - how many hours a week? <1/1-2/2-3/3-4/4> / 1-2-3-4-5
  - how many months a year: <1/1-3/3-4/4-5 > 1-2-3-4-5
10. In comparison with others of my own age I think my physical activity during leisure time is ... much less/less/the same/more/much more 1 - 2 - 3 - 4 - 5
11. During leisure time I sweat ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
12. During leisure time I play sport ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
13. During leisure time I watch television ... never/seldom/sometimes/often/very often 5 - 4 - 3 - 2 - 1
14. During leisure time I walk ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
15. During leisure time I cycle ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
16. How many minutes per day do you walk and/or cycle to and from work, school and shopping? <5/5-15/15-30/30-45/45> / 1-2-3-4-5
*17. During leisure time I do do-it-yourself activities ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
*18. During leisure time I work in the garden ... never/seldom/sometimes/often/very often 1 - 2 - 3 - 4 - 5
*19. How many hours per day do you sleep on average? <5/5-6/6-7/7-8/8-9 5 - 4 - 3 - 2 - 1
* Questions 17-19 were added to the original questionnaire.

Calculation of the score for question 9

If type of sport = 1 then Intensity = 0.76
  3 1.26
  5 1.76

If number of hours a week = <1 then Time = 0.5
  1-2 1.5
  2-3 2.5
  3-4 3.5
  >4 4.5

If number of months a year = <1 then Proportion = 0.04
  1-3 0.17
  4-6 0.42
  7-9 0.67
  >9 0.92

Question 9 = (Intensity, Time, Proportion) + (Intensity, Time, Proportion) = 0/0.01 <4/4-<8/8-<12/12
* 0 is given to people who do not play sport.

Calculation of the indices of physical activity

WORK INDEX = (11 + 12 + 13 + 14 + 15 + 16 + 17 + 18)/8
SPORTS INDEX = (19 + 110 + 111 + 112)/4
LEISURE TIME INDEX = (113 + 114 + 115 + 116 + 117 + 118 + 119)/7
TOTAL INDEX = work index + sports index + leisure time index