

## Paper 3

### How methods affect measures of physical activity in 8 European countries: The Eupass Experience

Randy Rzewnicki, Ilse De Bourdeaudhuij, Timo Stahl, John Welshman, YvesVanden Auweele, Heiko Ziemainz, Anne Vuillemin, Alfred Rütten, Wim TM Ooijendijk, Maarten Stiggelbout, and Federico Schena

Submitted *Social and Preventive Medicine*

#### ABSTRACT

*Objectives* : To compare designs and survey modes for physical activity (PA) data collection in the European Physical Activity Surveillance System (EUPASS) project. We investigate which of three selected methods (design/mode composites) was best in terms of response rates and sampling of the population and compare differences in reported minutes of PA of various intensities between methods.

*Designs*: cross sections, panel and continuous. Three datasets collected at the same time in 8 countries were compared. The survey techniques are mailed questionnaires and telephone interviews.

*Subjects*: Random samples of European adults from 8 countries (total N=2460)

*Results* : Mail method provided lower response rates than the telephone methods in the total sample and in 7 of 8 countries. The process of repeated contacts for the Panel method produced the lowest response rates of all methods in 7 of 8 countries. In most countries, the samples differ on gender and age from the expected values and vary between methods. Males were over-represented in mail samples and females were over-represented in telephone samples. In general the oldest and the youngest age groups were under-represented. Analyses of the total sample found differences between methods in moderate PA and sitting. Panel samples had significantly higher mean scores than the continuous and mail samples in several PA categories.

*Conclusions* : The low response rates and higher reports of PA from the Panel method contrast with the better profile of the Continuous method. Standardization of survey procedures is recommended.

#### INTRODUCTION

Epidemiological and experimental evidence has demonstrated a strong relationship between a sedentary lifestyle and risk of cardiovascular and other diseases as well as reduced well-being (USDHHS, 1996) . Harmonization and standardization of methods for collecting

physical activity (PA) data have yet to be achieved and empirical evidence supporting the efficiency of one survey method or another is in short supply. The European Union (EU), in order to provide comparable health information at EU level, has begun to explore the development of common procedures and protocols for measuring PA indicators in different EU countries. The European Physical Activity Surveillance System (EUPASS) project, reported here and elsewhere in this issue, was designed to contribute to that process.

Three modes are widely used for PA surveys of populations: personal interviews delivered face-to-face or by telephone, and self administered questionnaires delivered by mail. The three data collection designs commonly used for PA surveys are cross-sections, continuous (also called time series) and panel (also called repeated measures). Cross-sections provide data on samples at a single moment. The Panel design focuses on changes in the behavior of individuals in the variables of interest over time by collecting data from the same individuals on repeated occasions. Continuous data collection attends to population level trends by collecting data from new samples on a regular time schedule, such as monthly throughout a year (McQueen, 1999). To our knowledge, no studies have collected PA self reports in different countries using the same instrument and compared different survey methods<sup>1</sup>.

The purpose of this study was to compare methods for the collection of PA self reports in eight EU countries in the same time period. The first aim was to investigate how methods affected sampling in each country. For this, we examine the response rates and sample characteristics. The second aim was to investigate differences in the reported minutes of physical activity of different intensities between the three research methods.

The design of the study is described in more detail elsewhere (Ruetten, Ziemainz, Schena et al., 2003; Ruetten, Vuillemin, Ooijendijk et al., 2003), and will be summarized here. Three surveys were conducted in 8 countries. A panel study and a continuous study, both using telephone interviews, commenced simultaneously (Panel 1, Continuous 1). The panel study collected repeatability data after 1-3 weeks (Panel 2), (see results in Ruetten, Ziemainz, Schena et al., 2003) and a third data collection after 6 months (Panel 3). The continuous study collected telephone interview data during each of the six months (C1-C6) from new respondents each time. The third method was a mail survey which was conducted at the same time period as P3 and C6. As the purpose of this article is to compare the three methods used, we examine the P3, C6 and Mail data samples to avoid seasonal effects on PA.

---

<sup>1</sup> Method, as used here, refers to the combination of mode of delivery and design.

## **METHODS**

### **Samples**

In total 2402 adult respondents from Belgium (B), Finland (FIN), France (F), Germany (D), Italy (I), Netherlands (NL), Spain (E) and the United Kingdom (UK) completed the three survey protocols (Panel 3, Continuous 6, Mail) in Autumn 2000. Table 1 displays selected characteristics of 23 different samples and comparisons with standard national reference values.

### **Measures**

For the measures reported here, in each country the survey protocols were identical in content and question order across methods. Between countries only language differed. The International PA Questionnaire (IPAQ, short last 7 days version) contains 9 questions which assess frequency, intensity and duration of vigorous and moderate PA, walking and sitting. The survey items and results are described elsewhere (Ruetten, Ziemainz, Schena et al., 2003; Ruetten, Vuillemin, Ooijendijk et al., 2003; Ruetten, Rzewnicki, Ziemainz, et al, 2003).

### **Procedure**

Different sampling designs were used, but all were random samples with no stratification or selection criterion except age (18+). Sources of the telephone samples included a population register (FIN), Random Digit Dialing (RDD) (B, D, UK), and lists obtained from telephone companies (F, I, E, NL). If the preliminary source being sampled was a telephone number or a street address, then within the household the 'last birthday method' was used to select individuals. Samples for the Mail method were obtained from commercial sources (e.g. marketing agencies) except in Finland (population register) and France (telephone list).

Two countries announced surveys: Finland sent pre-letters explaining the telephone interview. France requested cooperation with the mail survey by telephone. For the telephone surveys, two countries used a paper questionnaire (E, UK). The others used computer-assisted telephone interviewing software, often referred to as CATI. CATI manages the dialing of telephone numbers and displays a EUPASS questionnaire on the computer screen during each interview. For the Panel study, most countries asked respondents at the end of the first interview for cooperation with the follow ups. Two (FIN, UK) asked at the beginning of the interview.

Insert table 1 here

Insert table 1 here

Insert table 1 here

For the translation of the survey into the language spoken in each country, a standard back-translation procedure was used. It began with translations from the English original by official translators who were native speakers of the target language. The translated texts were each vetted by separate focus groups of bilingual adults in each country (e.g. French-English or Italian-English) who were similar to the target population. After making the recommended changes, two further steps were made with the improved versions. It was translated back into English by a native English speaker who was expert in the field, but unrelated to the research project. The results of the “back-translation” were sufficiently close to the meaning of the original English text to be assured that the documents were comparable. The improved text was also presented to a second group of bilingual people similar to the target population, who evaluated it and suggested improvements.

The EUPASS survey, including the IPAQ, was administered via telephone by trained interviewers. Interviewers were usually university students or recent graduates. Experience levels varied widely, from zero previous interviews (F, E, I, UK ) to over 2000 (FIN). Half of the countries (B, D, FIN, I) had formal monitoring and feedback procedures. The others did not monitor interviewers after training.

Interviewers used the telephone script protocol described by the IPAQ standardization group (Craig, Marshall, Sjöström et al, in press). In most countries calls were not made during working hours or on weekends. The standardized interview took 10 to 20 minutes, and in countries using a CATI system, the data were entered directly into the computer. The mail protocol was an abbreviated Dillman total design method (Dillman 1978) , and included a cover letter separate from the questionnaire document, a post paid addressed return envelope and a reminder postcard.

The development of the survey content and ordering of items was centralized at project headquarters (HQ). The English version of the questionnaire document, telephone script and translation procedure were delivered to each country. Non-response criteria, categories and codes were created at HQ and distributed for use in all country reports. Data handling, cleaning, and statistical calculations of response reports and survey data were all centralized.

#### Data handling and Statistical analyses

Chi-square tests were used to examine differences in the response rates between the Panel, Continuous and Mail methods. For tests comparing sample characteristics with national data, chi squared statistics (with known expected values) were calculated to determine if the

observed results differed from the expected values. The reference data used for gender, age and employment status was considered internationally comparable as it was taken from standard sources (e.g. Eurostat yearbook 98/99). Age was categorized in decades (e.g. 25-34, 35-44 etc), to allow comparison with international data. For work status, only the employed and unemployed categories for individuals up to 64 years old were compared with national statistics (Eurostat yearbook 98/99). In the three countries where income data were available (FI median, NL tertiles, BE quartiles), comparison tests were conducted.

Data cleaning procedures for the IPAQ items were followed (Craig, Marshall, Sjöström et al, in press) and were supplemented by truncation at 10 hours per day per activity (e.g. maximum daily values of 10 vigorous hours, 10 moderate etc.). As all physical activity scores still showed a strong skewedness, and standard deviations were as much as double the mean scores, a logarithmic transformation approximated the normal distribution, a procedure used by other studies in the physical activity domain (De Backer, Kornitzer, Sobolski et al., 1981; De Bourdeaudhuij & Sallis, 2002).

To assess the robustness of the association between method and PA, we controlled for four mediating variables: gender, age, education, and job status. The question about job status was recoded to two categories as follows: 1) those in paid employment and students; 2) the unemployed, homemakers, retirees and others.

Univariate analyses of variance were executed with the log transformed activity minutes of different intensities as dependent measures, with method as group factor (continuous, panel, mail), and with gender, age, years of education and work status as covariates.  $P < 0.05$  was considered to be significant.

## RESULTS

### Response rates

Table 2 shows the response rates from the data collected at the same time period (Continuous 6, Panel 3, Mail). For the total samples, the methods differed significantly from one another. Continuous 6 (C6) produced the highest and Panel 3 (P3) the lowest response. The same pattern of response appears in most countries. All differences between response rates are significant ( $p < 0.001$ ) except in Finland, where no differences were found. An other pattern is also visible in Spain, where Mail is lower than Panel 3 and France where the Mail is higher than both telephone methods. The response rates for C6 are consistent with the other continuous samples (C1-C5 data not shown) except that C6 rates are higher than average for UK (mean C1-6=25.5%) and Finland (mean C1-6=54.5%). Response rates for both telephone methods at first contact (P1 & C1-6) (see Ruetten, Ziemainz, Schena et al.,) did

not differ significantly in the total sample nor in 7 of 8 countries.

## Chapter 2: METHODS & MEASUREMENTS

Table 2 : Effective sample sizes (response rates, %) and chi-square tests of differences between response rates.

	<b>Continue 6</b>	<b>Panel 3</b>	<b>Mail</b>	<b>Chi-square</b>
<b>Belgium</b>	101 (40.9)	79 (12.	206 (35.0)	108.9***
<b>Finland</b>	101 (62.3)	112 (51.6)	134 (58.3)	4.6
<b>France</b>	99 (24.0)	67 (13	131 (52.4)	128.3***
<b>Germany</b>	117 (51.1)	145 (15.2)	67 (19.1)	142.7***
<b>Italy</b>	100 (31.2)	91(14.5	148 (29.6)	45.7***
<b>Netherlands</b>	101 (43.0)	76 (23.	108 (25.4)	29.7***
<b>Spain</b>	100 (44.1)	100 (36.2)	22 (7.3)	103.0***
<b>United Kingdom</b>	106 (33.9)	79 (14.	70 (18.6)	46.9***
<b>ALL</b>	<b>825 (38.4)</b>	<b>749 (18.6)</b>	<b>886 (29.3)</b>	<b>295.6***</b>

\*\*\* p<0.001

### Representativeness

Summarizing, gender was misrepresented in half of the telephone samples, and in most mail samples. In the telephone samples, women were over-represented in 7 of the 8 cases where differences were observed. In the mail sample, men were over-represented in 4 of 5 cases. The age data differed significantly from expected values in the majority of samples. In general the oldest and youngest age groups were under-represented.

### IPAQ

In Table 3 the untransformed means and standard deviations of activity minutes and

### F-values

for tests of the transformed data are reported showing some differences between research methods for reported minutes of physical activity of different intensities.

Table 3 : Raw Means (SD) for minutes of PA per week for the combined data from 8 countries and F-values comparing methods after logarithmic transformation. (All Walking = vigorous, moderate & slow pace walking; TOTAL minutes = vigorous PA + moderate PA + vigorous and moderate walking)

<b>Minutes per week</b>	<b>Continuous-6 Mean (SD)</b>	<b>Panel-3 Mean (SD)</b>	<b>Mail Mean (SD)</b>	<b>F-test of log scores</b>
Vigorous PA	238 (587)	239 (499)	230 (464)	2.55
Moderate PA	266 (500)	325 (550)	346 (586)	3.58*
All Walking	618 (846)	589 (765)	628 (842)	0.72
Sitting	2194 (1064)	2370 (1125)	2506 (1076)	17.11***

---

TOTAL minutes	999 (1266)	1018 (1207)	960 (1221)	3.97*
---------------	------------	-------------	------------	-------

---

\* p<0.05; \*\*\* p<0.001

Overall analyses for the total research sample showed significant differences between methods for activities of moderate intensity ( $p<0.05$ ), for total amount of physical activity of at least moderate intensity ( $p<0.05$ ), and for sitting ( $p<0.001$ ). For all three analyses the highest means were found for the panel method. Post hoc tests showed that the panel method resulted in significantly higher mean scores than the continuous method and the mail method for total amount of physical activity, in significantly higher mean scores than the continuous method for activities of moderate intensity, and in significantly higher means scores than the mail survey for sitting. A note to the reader, the differences reported here between methods cannot to be seen with the data appearing in Table 3, as the tests were conducted on the logarithmically transformed data.

In general these differences are only minimally reflected in the analyses within countries, probably due to the smaller sample sizes. Only in France was the same trend found, with significantly higher reports of activities of moderate intensity ( $p<0.05$ ) and of minutes of sitting ( $p<0.001$ ) using the panel method. In the Netherlands, the panel and the continuous method also resulted in significantly higher scores on sitting minutes ( $p<0.001$ ) than the mail survey. However, in the Netherlands, significantly lower scores on moderate activity were found in the continuous method compared to the mail survey ( $p<0.001$ ). In Spain, where no mail data was tested, the panel method also resulted in higher scores for total minutes of at least moderate intensity activity ( $p<0.01$ ) compared with the continuous method. The opposite effect was found for sitting, with the panel method resulted in significantly lower sitting scores than the continuous method ( $p<0.05$ ).

Further, some non-systematic differences between research methods were found within countries for vigorous physical activity and for walking. In Belgium, the mail survey resulted in lower scores of vigorous activity compared to the panel method ( $p<0.05$ ), while in the United Kingdom the mail survey resulted in significantly higher scores on vigorous activity compared to the continuous method ( $p<0.01$ ). For walking, significantly higher scores were found in Germany with the mail survey than with the continuous and panel method ( $p<0.01$ ). In contrast, significantly lower scores were found with the mail survey for walking in the Netherlands compared to the continuous method ( $p<0.01$ ).

It can be concluded that the 3 methods produced significantly different response rates in most countries, and some differences in PA reports for all countries combined. Response

rates are generally low, highest for the Continuous method and lowest for the Panel. The Panel method produced higher reports of moderate PA than the Continuous method, higher reports of total PA than Mail and Continuous methods, and higher reports of sitting than the Mail method.

## **DISCUSSION**

### Response rates

Loss of subjects over time is a common problem in longitudinal research and in this study as well. The end product of the Panel method, as expected, shows the lowest response rates. The response rates for Panel 3 did not exceed 15% in 5 countries and are lower than the other methods for all countries except Spain. Only Finland, (51% at Time 3) could consider response rates from repeated telephoning of the same individuals acceptable for population surveys.

The procedures for the Continuous and the Panel methods were identical except for the request for future participation. Thus few if any differences were expected in response at initial contact and the fact that 7 of 8 countries had this result implies that these two telephone procedures were highly comparable.

Consistent with the literature (Streiner & Norman, 1995), the mail response rate was lower than the telephone methods (i.e. data from C1, P1, & Post ) for the total sample. At the national level, this was confirmed in all countries, except in France where the mail response was twice as much as the initial telephone response. In any case, a significant improvement in all response rates is required for PA monitoring. The relatively high response rate achieved with the mail method in France may be due to the pre-call. The French pre-call was a selection of randomly selected respondents before the questionnaire was sent, which means that the actual response rates are probably lower.

It is important to note that most countries did not use the same sources and procedures for their Mail and telephone samples, which likely explains some differences in the sample characteristics and response rate. Finland and France used the same sources for all samples, but then processed the data differently before arriving at a net sample. Other countries had different sample sources that may have contained idiosyncratic variations. For example, although the company supplying Belgian addresses claimed coverage of 94% of all residential addresses in Belgium, only a "head-of-household" was named and over 80% of these were male, which may explain the lower proportion of females sampled there.

Only Finland has response rates for both telephone methods above 50%. Why are most countries so low? First, in Finland telephone health surveys typically exceed 60% response (Hyypä & Maki, 2001), which is considered exceptionally high for Europe. Secondly, Eupass procedural differences may have contributed to differences between countries.

Variations in interviewer training and experience, supervision and CATI usage were not tested, thus their impact cannot be determined. However, interviewer training, monitoring and feedback are important for survey quality (Frey & Oishi, 1995) and their absence help explain some of the low response rates. In Finland Eupass researchers sampled from population registry, mailed a pre-letter, and called GSMs as well as fixed phones. Their highly experienced student interviewers used CATI, and received supervision and feedback. Whether it is any one or the sum of these procedures or other features specific to Finnish samples which produced the high response rates could be profitably investigated. We note that the absence of central population registers from which to sample citizens requires the use of RDD samples and thus excludes pre-letters. Therefore other procedures to enhance response rates should be considered.

#### Representativeness

Poor sample coverage results in limited generalizability due to the data not being representative of the population from which it was selected (Frey & Oishi, 1995). The EUPASS samples differ statistically from reference data, although very atypical gender and age characteristics is rare for most countries. The differences are probably due to the small sample sizes and the random sampling, with no quota controls.

#### IPAQ Data Skewed

Severe skewing of reports of PA at all intensities contrasted with the more normally distributed sitting data. It has been reported for other Eupass data (Ruetten, Ziemainz, Schena et al., 2003) that the medians for vigorous and moderate PA and walking were usually much smaller than the mean scores (i.e. medians are about half as large as the means). This indicates the majority of the respondents in each sample reported small amounts of PA and small proportions reported relatively extremely high amounts of PA, severely distorting the distribution. Although the implications for analyses and interpretation are far-reaching, discussion is beyond this report's scope.

#### IPAQ differences between Methods

Responses from people who are answering IPAQ telephone interviews for the third time (P3) show more PA than the people answering it for the first time (C6). This could be due to learning effects, higher motivation exhibited by people who submit to multiple interviews or other factors. The only procedural difference between P3 and C6 is the number of contacts. The two previous interviews could have taught the Panel respondents about the

research, especially the content, including time spent sitting, low and moderate intensity PA. In the same line, people answering the IPAQ telephone interview for the third time (P3) report they sit more hours than the people completing the written survey for the first time (Mail). Their increased attention to these behaviors (awareness-effect), as well as the Hawthorne effect (i.e. knowing they are part of research surveillance, Mayo, 1933) may have influenced their behavior. Whether the behavior influenced was PA or interview reporting is a topic for further research. Besides learning and motivation effects, the social interaction on the telephone, with higher time and social desirability pressures, may exert differential effects on PA reporting than completing a written protocol. In the future, interaction effects must be considered as we compare methods differing in mode (telephone vs. mail) and design (1st vs. 3rd completion).

The differences found in the total sample between methods on moderate and total PA and sitting, may be due to the relative difficulty respondents have in assessing and reporting vague activities. The moderate PA and walking components of the total PA figure, as well as sitting, are so commonplace that “reports” of such are actually estimates. These estimates are more likely to be influenced by various factors including learning, attention and motivation, which may have been higher in the Panel 3 respondents. In contrast, bouts of vigorous PA are discrete and salient events more easily recalled and reported.

The differences between methods on the IPAQ variables (moderate and total PA and sitting) could also be the result of differences between the samples on other determinants that were not included as covariates. For example income, health status, nationality, BMI, marital status, hours of work, type of job, and smoking are related to PA (Dishman 1993; Sallis & Owen 1999). Moreover, large differences between countries appeared on PA reporting with the different methods used. This means that more research within countries is necessary, using larger sample sizes, to investigate specific cultural trends within countries affecting PA reporting by telephone or mail.

Although it is beyond the scope of this paper to comment on the absolute levels of PA reporting, some information might be necessary to compare the three methods used. Preliminary results showed a tendency for the IPAQ questionnaire to over-report PA (Rzewnicki, Vanden Auweele & DeBourdeaudhuij, 2003). As reported elsewhere (Ruetten, Ziemainz, Schena et al., 2003) some evidence also exists that the EUPASS data also suffer such problem. From this, it could be argued that methods which lead to the least over-reporting (in our case Continuous data collection) have to be preferred above methods which result in highest reported levels of physical activity (Panel method).

### Limitations and shortcomings

This study may have been constrained by certain methodological limitations. All data are from self report, and self report of PA is subject to a variety of factors, most notably over-reporting of time and intensity (Montoye, Kemper, Saris & Washburn, 1996). Social desirability is an important factor in responding to surveys, and we may speculate that our respondents considered it desirable to report high amounts of PA. Further, the higher status groups, which are quite large in most EUPASS samples, may have tended to over report PA since their peers value a healthy lifestyle more than individuals in lower classes (Droomers, Schrijvers, van de Mheem & Mackenbach, 1998).

In the literature, interviewer experience is associated with interpretation of the meaning of questions, errors in skip patterns (i.e. questions or instructions related to items that can be skipped) and communicating the answers that they want to hear (e.g. Streiner & Norman, 1995). However, such general shortcomings can be prevented by adequate training, follow-up, supervision and feedback to interviewers. A related question is whether experienced interviewers have enough knowledge about the content of PA surveys. The IPAQ telephone questionnaire demands respondents count minutes of activity in their head, and report this very quickly to a stranger who may have expectations that high reports of PA is good. Moreover, the telephone interview started with questions about vigorous activity, which makes it very hard for interviewers who are not experienced in controlling over-reporting of PA to intervene adequately. It is possible that university students with few interviewing skills but who are more knowledgeable of the content of PA surveys and of the problem of over-reporting could produce better data than more experienced interviewers with less knowledge of the content.

### Conclusions

Considering the total sample of data from 8 EU countries, conclusions are tentative. The Panel method produces a select sample. As the Panel method resulted in significantly higher reports of IPAQ components than the other methods, and the Panel resulted in the absolute lowest response rates in every country, we suggest it is less appropriate for population surveys than the Continuous method. The Mail method produced the lowest responses at first contact, and had more male than female respondents. The Mail method may reach segments of the EU population who are not amenable to telephone surveys, and despite its shortcomings, we see some utility. The Continuous method has fewer shortcomings than the other methods

and thus is recommended.

Considering the difficulties of collecting data in many different countries, a healthy skepticism of pan-European survey data is reasonable, avoiding unduly high expectations. Standardization of survey procedures following the best practices is recommended due to the consistent high quality and quantity of data that can be produced. Where it is possible, sampling from population registers and sending pre-letters has advantages. Also using interviewers who are experienced with PA surveying, as well as survey administration which includes highly qualified supervision and CATI is to be recommended.

### **Acknowledgements**

EUPASS was funded by the Health Monitoring Program of the European Union (European Commission, Luxembourg, agreement number VS1999/5133 (99CVF3-502), A. Rütten, project leader). Also the authors wish to acknowledge the support of the following individuals for their contributions to the realisation of this research T. von Lengerke, M. Sjoestroem, P. Godin, M. Lejeune, K. Schmitt, A. Haase, L. Kannas, P. Ciddo, JM Oppert, N. Rodriguez Avila, J. Rodriguez Diaz, J. Vinck, A. Gatrell, D. McQueen.

## References

*Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis J F, Oja P*, and the IPAQ Consensus Group and the IPAQ Reliability and Validity Study Group International Physical Activity Questionnaire (IPAQ): 12-country reliability and validity. *Med Sci Sports Exerc*; in press.

*De Backer G, Kornitzer M Sobolski J, Dramaix M, Degre S, DeMarneffe M, Denolin H* (1981). Physical activity and physical fitness levels of Belgian males aged 40-55 years. *Cardiology* 67; 110-128.

*De Bourdeaudhuij I, Sallis J* (2002) Relative contribution of psychological determinants to the prediction of physical activity in three population based samples. *Prev Med* 34; 279-288.

*Dillman DA* (1978). *The Total Design Method* New York, New York: John Wiley and Sons.

*Dishman RK* (1993). Exercise adherence, in Singer RN; Murphey M; Tennant LK (eds), *Handbook of research in sport psychology*. New York: Macmillan, 779-798.

*Droomers M, Schrijvers CTM, van de Mheem H, Mackenbach JP* (1998). Educational differences in Leisure time physical inactivity: a descriptive and explanatory study. *Soc Sci Med & Health* 47 (11); 1665-76.

*Folsom AR, Cook TC, Sprafka JM, Burke GL, Norsted SW, Jacobs DR* (1991). Differences in leisure-time physical activity levels between blacks and whites in population-based samples: the Minnesota Heart Survey. *J Behav Med* 14; 1-9.

*Frey JH, Oishi SM* (1995). *How to conduct interviews by telephone and in person* Sage: London.

*Hyypä MT, Maki J* (2001) Individual-level relationships between social capital and self-rated health in a bilingual community. *Prev Med* 32; 148-155.

*Mayo E* (1933). *The human problems of an industrial civilization*. New York: MacMillan.

*McQueen D* (1999). A world behaving badly: The global challenge for behavioral surveillance. *Am J Public Health* 89;1312-14.

*Montoye HJ, Kemper HGC, Saris WMH, Washburn RA* (1996). *Measuring PA and Energy Expenditure*. Champaign IL: Human Kinetics.

*Sallis JF, Owen N* (1999) *Physical Activity &*

*Behavioral medicine*. London: Sage.

*Rütten A, Rzewnicki R, Ziemainz HR, Schena F, Schmitt K, Stiggelbout M, Ooiendijk WTM, Stahl T, Vanden Auweele Y, Vuillemin A, Welchman J*. (2003) Towards a European health monitoring system. Results of a pilot study on physical activity. In *McQueen D, Puska P*, editors. *Global behavioral risk factor surveillance*. New York: Kluwer, In press.

*Rütten A, Vuillemin A, Ooijendijk WTM, Schena F, Sjöström M, Stahl T, Vanden Auweele Y, Welshman J, Ziemainz H* (2003). Physical Activity Monitoring in Europe: The EUPASS Approach and Indicator Testing, *Pub Health Nutr* 6 (4); 377-384.

*Rütten A, Ziemainz H, Schena F, Stahl T, Stiggelbout M, Vanden Auweele Y, Vuillemin A, Welshman J* (2003). Using Different Physical Activity Measurements in 8 European Countries: Results of the EUPASS time series survey *Pub Health Nutr*; 6 (4); 371-376.

*Rzewnicki R, Vanden Auweele Y, DeBourdeaudhuij I* (2003). Addressing over reporting on the International Physical Activity Questionnaire (IPAQ) Telephone Survey with a Population Sample. *Pub Health Nutr* 6 (3); 299-305.

*Streiner DL, Norman GR* (1995). *Health Measurement scales: A practical guide to their development and Use* Oxford Univ Press: Oxford.

*USDHHS - US Department of Health and Human Services* (1996). *Physical Activity and health: A report of the Surgeon General Centers for Disease Control and Prevention*: Atlanta, USA.



