Martina Schmidt
Dr. sc. hum.

Design and evaluation of a short questionnaire for the assessment of physical activity in women

Geboren am 22. April 1967 in Frankfurt am Main
Diplom der Fachrichtung Mathematik am 22. Juni 1995 an der Universität Mainz

Promotionsfach: DKFZ
Doktorvater: Prof. Dr. sc. math. J. Wahrendorf

Physical activity (PA) is discussed as a preventive factor for many chronic diseases, and awareness is increasing that it needs to be included as a covariate in many epidemiological studies. Yet, measuring PA is difficult due to the complex nature of this variable. For large studies and for the retrospective assessment of PA, questionnaires are the method of choice. Due to frequently long latency periods for many diseases, long-term PA is of greater interest than current PA. Yet, there is lack of validated questionnaires assessing PA of distant age periods and comprehending all types of activity. Therefore, within the scope of this dissertation a short retrospective questionnaire for total physical activity of a typical 24h-day was designed. This 24h-questionnaire, in combination with special interviewing techniques, was evaluated with respect to validity and reliability of the assessment of physical activity in distant age periods.

To the author's knowledge, this was the first study validating a questionnaire for physical activity in the distant past that covers all types of activity such as occupational, household, leisure, and transportational activities. Assessing full validity of this 24h-questionnaire would have required a cohort with detailed information on true total PA in the distant past. As such a cohort was not available, relative validity was evaluated by comparing the 24h-questionnaire with a differently structured PA questionnaire. Therefore, from an ongoing study on postmenopausal breast cancer 110 cases and 101 controls were randomly selected. The 24h-questionnaire was administered in a telephone interview more than two months after the main study interview, which included a detailed questionnaire on PA performed in the age periods 30 to 49 years and 50+ years. Total PA scores were derived from both interviews as metabolic equivalent (MET)-hours per week.

Special emphasis was put on statistical methods for the evaluation of questionnaires, because reviewing the literature revealed a discrepancy between theory and practice of questionnaire evaluations. A systematic literature research was performed on validation studies published within the last five years. It showed, that most commonly correlation coefficients were used (in 89.1% of studies), while Bland-Altman methods were included in only 21.7% of the publications. Both methods were compared by own simulations, and for all applied methods strengths and limitations were investigated. It was demonstrated that correlation coefficients may result in misleading conclusions. Thus, they should only be used with extreme caution and correct interpretation. Moreover, the

simulations and investigations revealed that Bland-Altman methods are the preferable approach for questionnaire evaluations. Bland-Altman plots should be presented together with some measures of dispersion of the individual differences between the questionnaire and the reference measure, e.g. the mean and standard deviation of the differences, or the 95% limits of agreement. Those methods enable to investigate systematic bias, random errors, outliers, trends, and describe the amount of disagreement between the compared measurements. As further was detected by own simulations, Bland-Altman plots should be interpreted carefully when it is known or presumed that the compared measures have markedly different reliability. Inspections of the scatter plots should go along with the interpretation of the Bland-Altman plots.

Therefore, our own evaluation studies were primarily based on Bland-Altman methods. Furthermore, potential sources of systematic and random errors were investigated. After adjusting the reference values to reflect 24 hours, the mean difference between both questionnaires was 3 MET*h/wk for the age period 30 to 49 years, with 53.6% of absolute differences below 35 MET*h/wk, i.e. showing good agreement. Further 28.9% of differences could be considered acceptable agreement. Differences between questionnaires were nearly identical for cases and controls, thus measurement errors seem to be non-differential with respect to cancer status. Results for the age period 50+ years were similar.

In addition, a separate repeatability study with 30 women indicated good repeatability. However, results need to be interpreted with caution, due to the small, non-representative sample.

The evaluated 24h-questionnaire appears to be a valid tool to distinguish between high and low levels of women's physical activity in the distant past. A more intense interviewer training on the assignment of light and moderate activities could enhance the validity. To further improve the questionnaire, separate reporting of each major activity pattern during the considered time frame might be enabled, instead of assessing an average pattern over the whole time period.

The 24h-questionnaire covers all activities of a day. It enables separate analyses by intensity level and analysis of inactivity, but not analyses by activity type. This questionnaire is especially useful for populations with many intermittent and unstructured tasks, as it is typical for many women. Cognitive interviewing methods were used to enhance the recall, still interview duration was only about 10 minutes for the assessment of PA in the distant past. Altogether, the 24h-questionnaire may be a useful instrument for studies on the effect of total physical activity at different age periods on various diseases, as well as for studies where physical activity may be an important covariate.