

RELIABILITY AND VALIDITY OF THE INTERNATIONAL PHYSICAL ACTIVITY QUESTIONNAIRE (IPAQ)

Dear Editor-in-Chief:

In the August 2003 issue of *Medicine & Science in Sports & Exercise*[®], Craig et al. (2) published an important article on the validation of the International Physical Activity Questionnaire (IPAQ). Lack of comparability was always a major limitation among studies aimed at quantifying physical activity (PA) levels (4,5), and therefore their study makes an important contribution. The authors compared IPAQ short and long versions, and each of these against a reference method, the CSA (now MTI) accelerometer. Their study (2) included 14 centers in 12 countries. The authors used Spearman's correlation coefficients (r_s) to show that (a) IPAQ reliability was very good ($r_s = 0.8$); (b) the short and long versions produce comparable results ($r_s = 0.67$); and (c) the criterion validity ($r_s = 0.30$) is, at least, comparable to most other self-report questionnaires (2).

However, as Bland and Altman (1) have pointed out, the correlation coefficient is an inefficient tool (a) to assess agreement between two instruments that provide continuous scores and (b) to evaluate reliability. Although this coefficient indicates how close the relationship between two variables is to a straight line, it gives no indication of the slope of this line (1,3). Systematic differences between instruments do not affect the correlation coefficient but may substantially affect agreement.

For example, we have compared the short and long versions of IPAQ in a southern Brazilian city, and the correlation coefficient between the methods was 0.61 ($P < 0.001$), close to the combined value found in Craig's study. However, using the Bland and Altman method, we concluded that the agreement between these methods was poor because: (a) the limits of agreement were extremely wide; (b) the average difference between the two methods was 69 min, almost half of the cut-off value of 150 min; and (c) the scores were consistently higher according to the long version. Therefore, the short version systematically underestimated PA levels.

Furthermore, Craig et al. (2) calculated concurrent validity based on the percent agreement between the two versions coded as dichotomous variables (sufficiently active or not). Agreement was high—at least 70% in nearly all centers. However, a certain degree of agreement is to be expected purely by chance, and this is why the kappa coefficient, not the percentage of agreement, is the statistic of choice, excluding chance agreement (3).

For example, in our study mentioned above, the agreement coefficient was 79%, but the kappa value was only 54%. In addition, the prevalences of physical inactivity according to the short and long IPAQ versions were, respectively, 42 and 28% ($P < 0.001$). This means that, despite an agreement of almost 80% and a correlation coefficient of 0.61, the short version of IPAQ overestimated the prevalence of inactivity by 50%. We suggest that Craig et al. (2) might wish to reanalyze their excellent database, to calculate the statistics we have proposed above.

Pedro Curi Hallal
Cesar Gomes Victora
Federal University of Pelotas
Pelotas, Brazil

DOI: 10.1249/01.MSS.0000117161.66394.07

REFERENCES

1. BLAND, J. M., and D. G. ALTMAN. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310, 1986.
2. CRAIG, C. L., A. L. MARSHALL, M. SJOSTROM, et al. International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Med. Sci. Sports Exerc.* 35:1381–1395, 2003.
3. FLEISS, J. L. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, 211–37, 1988.
4. KRISKA, A. M., and C. J. CASPERSEN. Introduction to collection of physical activity questionnaires. *Med. Sci. Sports Exerc.* 29:S5–S9, 1997.
5. LAPORTE, R. E., H. J. MONTOYE, and C. J. CASPERSEN. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep.* 100:131–146, 1985.