Before the development of the International Physical Activity Questionnaire (IPAQ), population measures of physical activity were country-specific, noncomparable, and mostly developed to assess leisure-time activity. Given the global increases in noncommunicable disease the need for internationally comparable physical activity surveillance measures was identified. An initial meeting at World Health Organization Headquarters in 1998 convened a group of physical activity scientists to plan the development and testing of such a measure, resulting in IPAQ. The purpose of this commentary is to reflect upon the first decade of experience with IPAQ, compare its intended to its actual use, and comment on its strengths and weaknesses as an addition to the armamentarium of self-report physical activity measures.

IPAQ development was premised on the need to develop international population measurement to assess 'total physical activity' across the domains of work, domestic tasks, active transport, and leisure time, because patterns of activity across domains were expected to vary widely by country. IPAQ was developed because of the desire for cross-country comparison and international physical activity surveillance. To enhance use across countries, the measures were designed to allow for country-specific examples of activities, and specific guidance was provided on adaptation and translation processes.

Two formats of IPAQ were developed. The short form IPAQ was comprised of 7 items covering all domains of activity within generic intensity-specific questionnaire items (IPAQ short). The long form IPAQ of 27 items measured domain-specific activity for each intensity-specific category. The 12-country evaluation of IPAQ demonstrated the instrument was as reliable and valid as other self-report physical activity questionnaires, that the short form was suitable for population surveillance purposes, and the domain-specific long form was suitable for research activities. The IPAQ short form provides numerous advantages to users seeking to establish population surveillance of physical activity, including its brevity and the provision of an overall total physical activity estimate. This instrument was designed primarily as a public health surveillance indicator, for large scale representative surveys, and to compare cross-national or regional estimates of physical activity. IPAQ exists in many languages and can be freely accessed at www.ipaq.ki.se and used without formal permission.

Prevalence estimates using IPAQ have now been published from several studies, with the International Prevalence Study of Physical Activity reporting on 20 countries, the 2002 World Health Survey reporting on 51 countries and Eurobarometer studies reporting on 15 European countries. Other studies examined the correlates of IPAQ-estimated physical activity among...
countries, including the physical environment, and a broad range of socioeconomic and demographic correlates using population samples from diverse countries.

One important category of evidence is the relationship between IPAQ values from epidemiological studies of physical activity and disease outcomes. IPAQ seems cross-sectionally related to obesity and other health risks in ways similar to other self report measures. There is limited longitudinal epidemiological evidence from IPAQ, but in one study 2172 cardiac patients were followed for 18 months. This study demonstrated the usual dose-response relationship of lower mortality and reduced CVD event rates predicted by higher IPAQ-assessed physical activity.

There are important limitations of IPAQ, both in its design and application in research studies. It was designed and tested in adults 18 to 69 years old but is often inappropriately used with younger adolescents or older adults. IPAQ short is limited to the assessment of ‘total physical activity,’ and cannot assess changes in any specific domain. For example, IPAQ use has occurred in small sample clinical studies, where the intervention mostly comprised more specific leisure-time PA advice, and IPAQ here may not be the best measure to detect program effects.

Although developed for population-level surveillance, IPAQ is sometimes used in intervention studies despite a lack of evidence to date that it is responsive to physical activity change. There are also concerns regarding the interpretation of the high prevalence estimates generated from the measure, such that at least three quarters of many populations meet or exceed 600 MET-mins, a level defined as ‘5 × 30 mins of ‘moderate’ (4 MET intensity) activity’ or equivalent. IPAQ provides very different prevalence rates compared with those derived from the myriad of leisure-time physical activity surveys. The ‘30 minutes × 5 days’ of moderate-intensity physical activity recommendation (the “public health recommendation”) is a relatively low threshold if “total daily activity” is assessed. This level is described as ‘moderate’ in the IPAQ scoring protocol because it would be achieved by most adults through background activity, such as work, housekeeping, and family care, that adults accumulate daily. The ‘high’ IPAQ category is recommended to account for this limitation in applying recommendations to self-reported ‘total activity.’

Despite extensive testing, there is a likelihood of misreporting with IPAQ due to challenges in recall and cross-cultural differences in reporting total physical activity. A concern for surveillance is the potential for differential measurement error between countries. Other measurement issues include likely question order effects, over reporting, and high variance, which means that small sample studies will be under-powered to detect between-group differences. High variance is compounded by using the IPAQ short form as a continuous measure, as its primary purpose was categorical reporting of PA levels as part of a surveillance system. Despite these caveats, the use of IPAQ has spread more rapidly than was expected and demonstrates the rapid adoption of a measurement innovation by many researchers, without sufficient consideration of its purpose and structure.

One important development following IPAQ was the testing of a related measure, providing an intermediary option between IPAQ short and long and that specifically provided domain-specific estimates. The Global Physical Activity Questionnaire (GPAQ), was developed using a similar process and is the recommended physical activity measure within the World Health Organization STEPiwise surveillance system. Using this domain-specific measure in prevalence studies is expected to be more policy relevant than short version of IPAQ estimates of vigorous, moderate, and walking activities, although at the expense of additional questionnaire length. Nonetheless, there are similar issues with respect to GPAQ use, its potential for measurement error and comparability among countries. An additional intermediate approach has used only the leisure and transport components of IPAQ long-form in Latin America (Hallal, this issue) as these two domains seem to be most relevant for public health interventions and surveillance.

Since the original development work there have been many studies replicating reliability and validity testing of IPAQ in very specific populations and minority subgroups. Overall this contributes relatively little new information to the understanding of IPAQ measurement properties. By contrast, research that is much needed, but less often carried out, is to explore cultural and language and social norm differences in question interpretation, to conduct cognitive testing of the IPAQ across countries, and to identify reasons why sometimes similar countries report IPAQ estimates that are substantially different.

In conclusion, IPAQ is one of two currently available short self-report or interview-based measures for international physical activity surveillance. It is the shortest measure available and should be used in large population studies with full understanding of its underlying assumptions and limitations. Problems remain with all self-report measures, but especially with IPAQ and GPAQ in the international arena: translations are variable, compliance with established protocols varies, and different centers use diverse methods for data cleaning and analysis. There is evidence that, for some between-country comparisons, these measures still have problems; but in most cases, IPAQ comparisons seem to be concordant with them and ranks countries similarly to other population studies. To minimize the concerns regarding use of self-report measures for surveillance, alternative approaches, including the use of objective population measurement for international comparison and surveillance should be pursued, especially in developing countries.
Author’s Note

The conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References


