

Reproducibilidad del tiempo en posición sedente evaluado con el *International Physical Activity Questionnaire* (IPAQ) y el *Global Physical Activity Questionnaire* (GPAQ)

Adriana Angarita Fonseca, Fisiot, Msc*

Diana Marina Camargo Lemos, Bact, Msc**

Myriam Oróstegui Arenas, Enferm, Msc*

Resumen

Objetivo: Evaluar la reproducibilidad y el nivel de acuerdo del tiempo en posición sedente evaluado con el IPAQ y el GPAQ. **Métodos:** Se realizó un estudio de evaluación de tecnologías diagnósticas. Los cuestionarios IPAQ y GPAQ fueron administrados por dos encuestadores a 92 adultos (42.2 ± 13.9 años), en dos oportunidades con un intervalo de tiempo entre 3 y 6 días, en el mismo orden de aplicación establecido aleatoriamente en la primera prueba. En el análisis, se evaluó la reproducibilidad prueba-reprueba de los ítems que miden el tiempo en posición sedente del IPAQ y del GPAQ y la reproducibilidad entre ítems, aplicando el Coeficiente de Correlación Intraclass (CCI (2.1)) y sus intervalos de confianza del 95% (IC95%). El nivel de acuerdo se estableció mediante el método Bland y Altman. **Resultados:** La reproducibilidad prueba-reprueba para el tiempo en posición sedente fue buena para el IPAQ (CCI: 0.77 IC95% 0.67; 0.84) y muy buena para el GPAQ (CCI: 0.83 IC95% 0.76; 0.89). El acuerdo fue pobre con un promedio de las diferencias de -0.04 h/d (límites de acuerdo: -4.95; 4.9 h/d) para el IPAQ y 0.15 h/d (límites de acuerdo: -4.2; 4.5 h/d) para el GPAQ. La reproducibilidad entre ítems fue muy buena en la primera prueba CCI: 0.81, (IC95% 0.73; 0.87) y segunda prueba CCI: 0.82 (IC95% 0.74; 0.88); el nivel de acuerdo entre ítems fue pobre para la primera -0.54 h/d (límites de acuerdo: -4.7; 3.6 h/d) y segunda prueba -0.6 h/d (límites de acuerdo: -4.9; 4.2 h/d). **Conclusiones:** La medición del comportamiento sedentario o hipoactividad física mediante los ítems del IPAQ y GPAQ presenta buena reproducibilidad pero pobre acuerdo, lo cual sugiere mejorar la medición de este comportamiento. Adicionalmente, los resultados mostraron que los ítems del IPAQ y el GPAQ proveen información similar. [Angarita A, Camargo DM, Oróstegui I. *Reproducibilidad del tiempo en posición sedente evaluado con el International Physical Activity Questionnaire (IPAQ) y el Global Physical Activity Questionnaire (GPAQ)*. MedUNAB 2010; 13:5-12].

Palabras clave: Cuestionario, reproducibilidad, sedentarismo, IPAQ, GPAQ.

Summary

Purpose: To evaluate the reliability and agreement of the time sitting items from the IPAQ and GPAQ. **Methods:** An evaluation of diagnostic technologies was done. The IPAQ and GPAQ were administered by two interviewers to a sample of 92 adults (42.4 ± 13.9 years), twice with an interval of time between 3 to 6 days, in the same order of application established randomly in the first test. In the analysis, we assessed test-retest reliability of the sitting items from the IPAQ and GPAQ and between items reliability by Intraclass Correlation Coefficient (CCI 2.1) and their confidence intervals at 95% (IC95%). The level of agreement was established by Bland-Altman method. **Results:** Test-retest reliability to time sitting items of the IPAQ was good (ICC: 0.77 (CI95%: 0.67-0.84) and very good to GPAQ (ICC: 0.83 CI95% 0.76-0.89). Agreement was poor with a mean difference of -0.04 (limits of agreement: -4.95, 4.9 h/d) to the IPAQ and 0.15 h/d (limits of agreement -4.2, 4.5 h/d) to the GPAQ. Between items reliability was very good in the first test (CCI: 0.81 CI95% 0.73-0.87) and in the second test (CCI: 0.82 CI95% 0.74-0.88); Agreement was poor in the first test with a mean difference of -0.54 (limits of agreement: -4.7; 3.6 h/d) and in the second test -0.36 (limits of agreement: -4.9; 4.2 h/d). **Conclusions:** The measurement of sedentary behavior or physical hypoactivity through the items of the IPAQ and GPAQ showed good reliability but poor agreement, it is suggested to improve the measurement of this behavior. Additionally, the results showed that the sitting item of the IPAQ and GPAQ provide similar information. [Angarita A, Camargo DM, Oróstegui I. *Reliability of time sitting items from the International Physical Activity Questionnaire (IPAQ) and Global Physical Activity Questionnaire (GPAQ)*. MedUNAB 2010; 13:5-12].

Key words: Questionnaire, reliability, sedentarism, IPAQ, GPAQ.

* Docente, Departamento de Salud Pública, Escuela de Medicina, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga, Colombia.

** Docente, Escuela de Fisioterapia, Facultad de Salud, Universidad Industrial de Santander, Bucaramanga, Colombia.

Correspondencia: Maestra Angarita, Departamento de Salud Pública, Escuela de Medicina, Facultad de Salud, Universidad Industrial de Santander, Carrera 32 N° 29-31, Bucaramanga, Colombia. E-mail: adriangarita@hotmail.com

Artículo recibido: 23 de febrero de 2010; aceptado: 23 de marzo de 2010.

Introducción

El sedentarismo ha sido asociado como factor de riesgo para las enfermedades cardiovasculares, la diabetes mellitus tipo 2, la osteoporosis, la obesidad y algunos tipos de cáncer, entre otros eventos en salud.¹ No obstante, la mayoría de estudios no ha medido o diferenciado el comportamiento sedentario de la actividad física baja o de la inactividad física. Esta dificultad, se puede explicar por el hecho de considerar que una persona sedentaria es aquella que es inactiva o que no es físicamente activa,^{2, 3} lo que resulta inadecuado, porque existe evidencia que indica que son comportamientos independientes con diferentes efectos sobre la salud.⁴

Biddle⁵ afirma que el sedentarismo es un comportamiento diferente caracterizado por bajo gasto energético, determinado más por ciertas conductas dominantes que por la ausencia de actividad física. Mientras que Pate *et al*⁶ se refieren al comportamiento sedentario o hipoactividad física como las actividades que no incrementan el gasto energético substancialmente sobre el nivel de reposo e incluye actividades como dormir, permanecer sentado o acostado, ver televisión y otras formas de entretenimiento basados en una pantalla.

Lo anterior señala que no hay consenso en su definición; adicionalmente, los autores no han medido consistentemente el comportamiento sedentario o hipoactividad física y mucho menos se cuenta con un instrumento válido y reproducible que mida este tipo de comportamiento, no solo en el tiempo libre sino en todos los dominios de la actividad física: en el trabajo, en el hogar, como desplazamiento y en el tiempo libre.

Clark *et al*⁶ revisaron ocho artículos que evaluaron la reproducibilidad prueba-reprueba de la medición del tiempo dedicado a ver televisión y otros comportamientos sedentarios fuera de la jornada laboral en adultos mediante diferentes cuestionarios; encontraron coeficientes de correlación intraclase (CCI) que oscilaron entre 0.32 y 0.93.

En un intento por estandarizar la medición de actividad física se desarrolló el Cuestionario Internacional de Actividad Física (*International Physical Activity Questionnaire*, por sus siglas en inglés IPAQ) en su versión larga y corta;⁷ los cuales fueron la base para que la Organización Mundial de la Salud (OMS) creara el Cuestionario Global de Actividad Física (*Global Physical Activity Questionnaire*, por sus siglas en inglés, GPAQ) que reúne las características del IPAQ corto y largo.⁸ Estos cuestionarios cuentan con un ítem que mide el comportamiento sedentario o la hipoactividad física, indagando sobre el tiempo que permanece sentada una persona en un día, ambos cuestionarios están dirigidos a adultos entre 18 y 65 años.

MacFarlane *et al*⁹ y Deng *et al*¹⁰ encontraron que la reproducibilidad de los METs-min/sem de hipoactividad física medidos con el IPAQ-corto oscilaron entre CCI 0.89 y 0.97, mientras que Kurtze *et al*⁹ y Papatthanasou *et al*¹² evaluaron la reproducibilidad de las horas/día de hipoactividad física, encontrando CCI entre 0.66 y 0.80.

En la literatura revisada no se encontraron estudios que evaluaran la reproducibilidad del tiempo en posición sedente evaluado mediante el GPAQ, a su vez no se han encontrado estudios publicados en la literatura científica, que evaluaran la reproducibilidad de los ítems de hipoactividad del IPAQ en población Colombiana.

Por lo tanto, el objetivo de este estudio fue evaluar la reproducibilidad prueba reprueba y el nivel de acuerdo de los ítems que evalúan hipoactividad física mediante el IPAQ y el GPAQ y la reproducibilidad y nivel de acuerdo entre ítems de estos cuestionarios en una población del área urbana de Bucaramanga.

Metodología

Se realizó un estudio de evaluación de tecnologías diagnósticas. La muestra fue seleccionada por conveniencia y estuvo conformada por adultos de ambos géneros con edad entre 19 y 70 años, residentes en Bucaramanga, quienes fueron evaluados en la Encuesta Basal del Programa Conjunto de Acciones para la Reducción Multifactorial de las Enfermedades No Transmisibles (CARMEN) realizada entre 2000 y 2001 y que participaron en el estudio Incidencia de Enfermedad Cardiovascular y sus Factores de Riesgo (INEFAC).

La recolección de la información se realizó entre septiembre de 2007 y enero de 2008 por dos fisioterapeutas previamente entrenadas, una con dos años de experiencia en la aplicación del IPAQ-corto y la otra sin experiencia. El análisis de la información se realizó en el 2009. Se midieron las variables género, estrato socioeconómico (bajo y medio-alto), estado civil (con o sin compañero permanente) y escolaridad (hasta primaria, secundaria y técnica-universitaria) para describir las características sociodemográficas de la población de estudio.

Medición del tiempo en posición sedente. Para evaluar el tiempo en posición sedente se usaron las traducciones oficiales disponibles para Colombia de la versión corta del IPAQ¹³ y la traducción oficial al español del GPAQ.¹⁴ En la tabla 1 se presenta el ítem que evaluó el tiempo dedicado a estar sentado para cada cuestionario: La unidad de medida del IPAQ fueron las horas en posición sedente en un día durante los últimos 7 días, mientras que para el GPAQ fueron las horas en posición sedente en un día típico.

Tabla 1. Medición de las horas en posición sedente según cuestionario

IPAQ	GPAQ
Ahora piense acerca del tiempo que Usted pasó sentado(a) en la semana durante los últimos 7 días. Incluya el tiempo en el trabajo, en la casa, estudiando y durante el tiempo de descanso. Esto puede incluir el tiempo que pasó sentado(a) en un escritorio, visitando amistades, leyendo, sentado(a) o acostado(a) viendo televisión. ¿Durante los últimos 7 días cuánto tiempo en total usted usualmente pasó sentado durante un día en la semana?	La siguiente pregunta se refiere al tiempo que suele pasar sentado o recostado en el trabajo, en casa, en los desplazamientos o con sus amigos. Se incluye el tiempo pasado ante una mesa de trabajo, sentado con los amigos, viajando en autobús, jugando a las cartas o viendo la televisión, pero no se incluye el tiempo pasado durmiendo. ¿Cuándo tiempo suele pasar sentado o recostado en un día típico?

Procedimiento. Se realizaron dos aplicaciones del IPAQ y dos del GPAQ en dos momentos diferentes mediante entrevista personal, con un intervalo de tiempo entre pruebas que osciló entre 3 y 6 días. En la primera prueba, que se realizó en las instalaciones de la Universidad Industrial de Santander, los participantes del estudio INEFAC respondieron una encuesta de 144 preguntas relacionadas con información sociodemográfica y factores de riesgo cardiovascular; inmediatamente, se les explicaron los objetivos de este estudio y fueron invitados a participar. Una vez firmado el consentimiento informado, se realizó la aleatorización del orden de aplicación de los dos cuestionarios, las encuestadoras registraron la información en forma consecutiva según la aleatorización previa; posteriormente, se concertó una cita en el transcurso de los siguientes 7 días para realizar la segunda prueba en la residencia del participante. Durante la segunda prueba, se aplicaron únicamente los dos cuestionarios de actividad física en el mismo orden de la primera prueba.

Análisis estadístico. Se aplicaron medidas de tendencia central y dispersión, medidas de posición o porcentajes según la naturaleza y distribución de las variables. Se evaluaron la normalidad de las variables mediante la prueba Shapiro Wilk y las características sociodemográficas por género aplicando las pruebas chi cuadrado y la prueba de Wilcoxon. Las diferencias entre mediciones fueron comparadas aplicando una prueba t pareada o la prueba de Wilcoxon.¹⁵ La reproducibilidad prueba-reprueba y entre ítems de los cuestionarios fue evaluada mediante el coeficiente de correlación intraclase (CCI (2.1))¹⁶ y el nivel de acuerdo entre mediciones del IPAQ y el GPAQ, así como entre ítems de los cuestionarios, se estableció aplicando la metodología de Bland y Altman.¹⁷

El CCI permite cuantificar la reproducibilidad entre diferentes mediciones de una variable cuantitativa medida en escala de razón; en donde se comparan dos o más mediciones por el mismo evaluador o dos o más mediciones con el mismo instrumento y puede usarse en muestras pequeñas.¹⁸ La forma del cálculo se basa en el modelo de análisis de varianza (ANOVA) con mediciones repetidas. La variabilidad total de las mediciones se puede descomponer en dos fuentes de variación, la que obedece a las diferencias entre los distintos sujetos y las diferencias entre las medidas para cada sujeto; esta a su vez depende de la variabilidad

entre las observaciones y una variabilidad residual o aleatoria asociada al error que conlleva toda medición.¹⁹ Por lo anterior, el CCI se define entonces como la proporción de variabilidad total que se debe a la variabilidad de los sujetos.^{16, 20} La interpretación del CCI se realizó mediante la clasificación de Altman: pobre (≤ 0.20); aceptable (0.21 – 0.40); moderada (0.41 – 0.60); buena (0.61 – 0.80); y muy buena (0.81 – 1.00) reproducibilidad.²¹

El análisis de Bland y Altman consiste en representar gráficamente las diferencias entre dos mediciones en el eje y frente a su media en el eje x, las cuales deben estar en las mismas unidades de medida. El intervalo de dos desviaciones estándar alrededor de la media de las diferencias incluye el 95% de las diferencias observadas. Este tipo de gráfico permite evaluar el error y el sesgo de las mediciones, identificar datos extremos, así como valorar tendencias. Existe un buen nivel de acuerdo cuando el promedio de las diferencias es cercano a cero, los límites de acuerdo son estrechos y el 95% de las observaciones está dentro de estos límites.¹⁸

La base de datos fue sometida a doble digitación y validada en el programa EpiInfo v 6.04d.²² Los diferentes cálculos estadísticos y el procesamiento de variables fueron realizados en el programa STATA/SE v. 9.1.²³ Se consideró un nivel $\alpha < 0.05$ para todas las pruebas.

Este estudio se consideró sin riesgo, según la Resolución 008430 del 4 de octubre de 1993 del Ministerio de Protección Social de Colombia²⁴ y fue aprobado por el Comité de Ética de la Universidad Industrial de Santander. Los participantes dieron su consentimiento informado escrito.

Resultados

Las características sociodemográficas se resumen en la tabla 2. Los 92 participantes mostraron una mediana de edad de 42.5 años (RIC 32.5 – 54.0) y el 66.3% correspondió al género femenino. Se detectaron diferencias por género, en cuanto al estrato socioeconómico que fue más alto para las mujeres ($p=0.02$) y el nivel de escolaridad, superior para los hombres ($p=0.04$).

Tabla 2. Características sociodemográficas de la muestra

Característica	Masculino (n=31)	Femenino (n=61)	Total (n=92)	Valor de p
Edad (Mediana y RIC)	38 (31.0-49.0)	43 (34.0-57.0)	42.5 (32.5-54.0)	0.165
Estrato socioeconómico				
Bajo	13 (41.9%)	12 (19.7%)	25 (27.2%)	0.023
Medio/alto	18 (58.1%)	49 (80.3%)	67 (72.8%)	
Compañero(a) permanente				
Si	20 (64.5%)	32 (52.5%)	52 (56.5%)	0.270
No	11 (35.5%)	29 (47.5%)	40 (43.5%)	
Escolaridad				
Hasta primaria	6 (19.4%)	26 (42.6%)	32 (34.8%)	0.040
Secundaria	18 (58.1%)	20 (32.8%)	38 (41.3%)	
Técnica/Universitaria	7 (22.6%)	15 (24.6%)	22 (23.9%)	

RIC: Rango intercuartílico (Percentil 25-Percentil 75)

Descripción del comportamiento sedentario. En la tabla 3 se observa que la mediana de tiempo que permanecen en posición sedente los participantes fue similar para la primera y segunda prueba del IPAQ y GPAQ. Aunque la mediana de horas por día en posición sedente fue mayor en el GPAQ comparado con el IPAQ-corto, diferencia que fue estadísticamente significativa en la primera prueba ($p=0.016$) pero no en la segunda ($p=0.142$). No se encontraron diferencias estadísticamente significativas por género.

Tabla 3. Distribución del tiempo en posición sedente según cuestionario y género (horas/día)*

Cuestionario	Género	Primera prueba	Segunda prueba
IPAQ-corto	Femenino	4 (2.2-6.0) ^a	4 (2.0-7.0)
	Masculino	5 (2.5-8.5)	5 (3.0-10.0)
	Ambos	4 (2.4-7.0)	4 (2.0-8.0)
GPAQ	Femenino	5 (2.5-7.0)	4 (2.0-7.0)
	Masculino	6 (2.5-10.0)	5 (3.0-10.0)
	Ambos	5 (2.5-8.0)	5 (2.0-8.0)

* Mediana (Rango intercuartílico)

Reproducibilidad prueba-reprueba y nivel de acuerdo de las mediciones del IPAQ y GPAQ. En la tabla 4 se observa entre buena y muy buena reproducibilidad prueba-reprueba para el tiempo en posición sedente para el IPAQ-corto y muy buena para el GPAQ, con CCI superiores a 0.73 para el primero y por encima de 0.82 para el segundo. Se observa que el GPAQ y la categoría masculino tienden a presentar coeficientes más altos; sin embargo, los CCI no presentan diferencias estadísticamente significativas por cuestionario y género, puesto que los intervalos de confianza no se cruzan.

Por otro lado, el promedio de las diferencias entre mediciones se distribuyó de manera normal en los hombres, mientras que para las mujeres y el total de la población la distribución fue no normal tanto para el IPAQ como para el GPAQ. El nivel de acuerdo de las mediciones de ambos cuestionarios muestra un promedio de las diferencias cercano a cero; no obstante, los límites de acuerdo oscilan entre -5.1 y 5.2 h/d mostrando un pobre nivel de acuerdo (figuras 1 y 2).

Tabla 4. Reproducibilidad prueba-reprueba y nivel de acuerdo de las mediciones del IPAQ-corto y GPAQ

Cuestionario	Género	Primera prueba ^a	Segunda prueba ^a	B & A ^b	CCI ^c (IC95%)
IPAQ-corto	Femenino	4.9 ± 3.3	4.8 ± 3.8	0.07 (-5.1; 5.2) ^e	0.73 (0.6; 0.8)
	Masculino	5.6 ± 3.7	5.8 ± 3.9	-0.3 (-4.7; 4.2) ^d	0.82 (0.7; 0.9)
	Ambos	5.1 ± 3.5	5.2 ± 3.8	-0.04 (-4.95; 4.9) ^e	0.77 (0.7; 0.8)
GPAQ	Femenino	5.2 ± 3.4	5.0 ± 4.0	0.18 (-4.2; 4.6) ^e	0.82 (0.7; 0.9)
	Masculino	6.5 ± 3.9	6.4 ± 3.8	0.07 (-4.1; 4.3) ^d	0.85 (0.7; 0.9)
	Ambos	5.7 ± 3.6	5.5 ± 4	0.15 (-4.2; 4.5) ^e	0.83 (0.8; 0.9)

^a Promedio y desviación estándar

^b Método de de Bland y Altman (se presentan el promedio de las diferencias y los límites de acuerdo del 95%)

^c Coeficiente de correlación intraclase (2.1)

^d Las diferencias entre mediciones muestran una distribución normal

^e Las diferencias entre mediciones no muestran una distribución normal

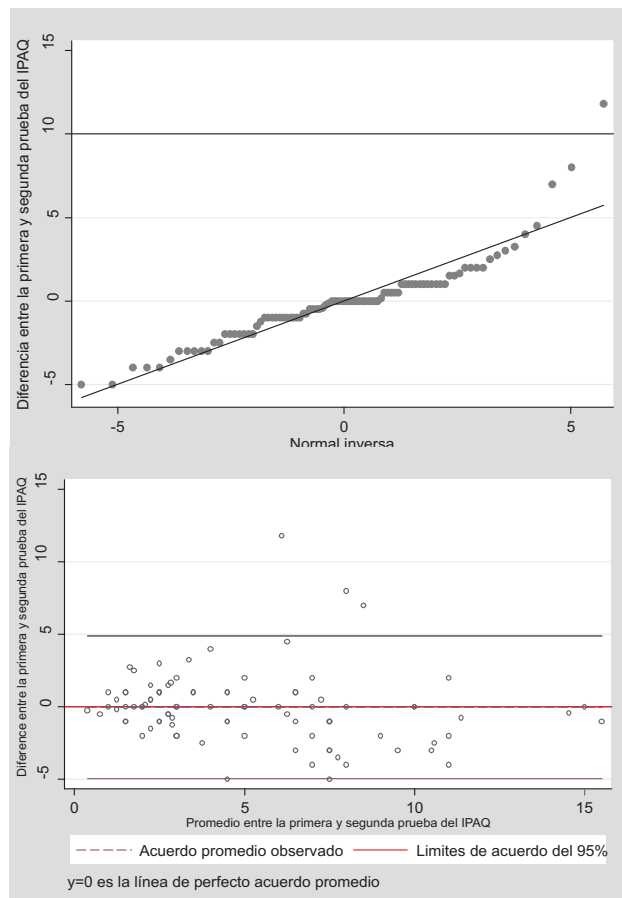


Figura 1. A. Gráfico cuantil normal de las diferencias entre la primera y segunda medición del tiempo en posición sedente del IPAQ. B. Gráfica de Bland y Altman entre la primera y segunda medición del tiempo en posición sedente del IPAQ (horas/día).

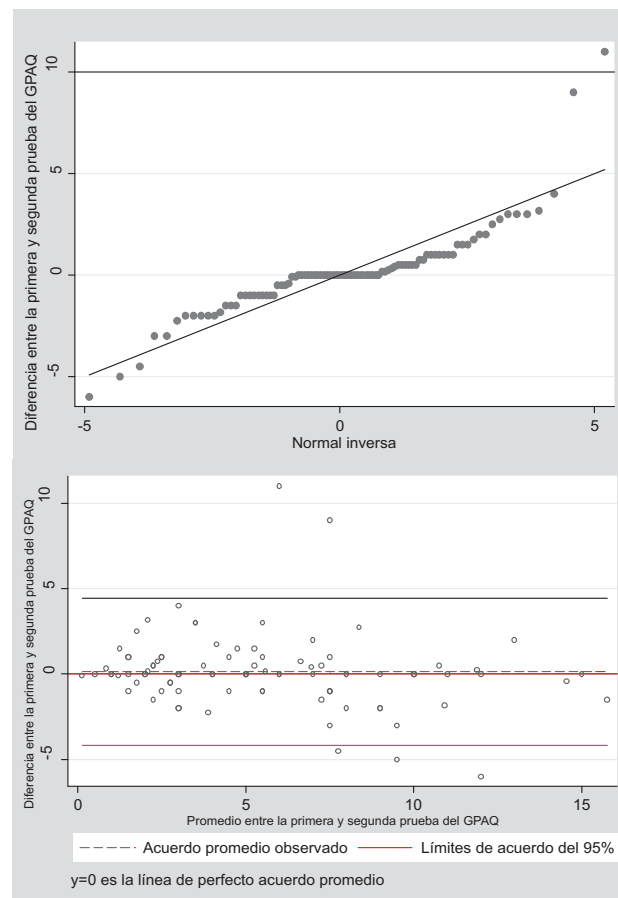


Figura 2. A. Gráfico cuantil normal de las diferencias entre la primera y segunda medición del tiempo en posición sedente del GPAQ. B. Gráfica de Bland y Altman entre la primera y segunda medición del tiempo en posición sedente del GPAQ (horas/día).

Reproducibilidad y nivel de acuerdo entre cuestionarios.

La reproducibilidad entre cuestionarios para el tiempo en posición sedente fue muy buena para el IPAQ-corto y el GPAQ (CCI>0.80). No se encontraron diferencias estadísticamente significativas por cuestionario y género en la reproducibilidad. No obstante, de forma similar a lo ocurrido en la reproducibilidad prueba-reprueba el nivel de

acuerdo entre cuestionarios fue pobre, en este caso los promedios de las diferencias estuvieron más alejados del cero, mostrando que el IPAQ-corto subestima el número de horas por día en posición sedente comparado con el GPAQ; adicionalmente, los límites de acuerdo fueron amplios oscilando entre -5.4 y 4.2 h/d (tabla 5). A su vez, el promedio de las diferencias entre ítems se distribuyó de manera

Tabla 5. Reproducibilidad y nivel de acuerdo entre cuestionarios en la primera y segunda prueba

Cuestionario	Género	IPAQ-corto ^a	GPAQ ^a	B & A ^b	CCI ^c (IC95%)
Primera prueba	Femenino	4.9 ± 3.3	5.2 ± 3.4	-0.35 (-4.5; 3.8) ^e	0.80 (0.7; 0.9)
	Masculino	5.6 ± 3.7	6.5 ± 3.9	-0.92 (-5.0; 3.2) ^d	0.83 (0.7; 0.9)
	Ambos	5.1 ± 3.5	5.7 ± 3.6	-0.54 (-4.7; 3.6) ^e	0.81 (0.7; 0.9)
Segunda prueba	Femenino	4.8 ± 3.8	5.0 ± 4.0	-0.24 (-4.7; 4.2) ^e	0.83 (0.7; 0.9)
	Masculino	5.8 ± 3.9	6.4 ± 3.8	-0.59 (-5.4; 4.2) ^d	0.80 (0.6; 0.9)
	Ambos	5.2 ± 3.8	5.5 ± 4	-0.36 (-4.9; 4.2) ^e	0.82 (0.7; 0.9)

^a Promedio y desviación estándar

^b Método de de Bland y Altman (se presentan el promedio de las diferencias y los límites de acuerdo de Bland y Altman del 95%)

^c Coeficiente de correlación intraclase (2.1)

^d Las diferencias entre mediciones muestran una distribución normal

^e Las diferencias entre mediciones no muestran una distribución normal

normal en los hombres, mientras que para las mujeres y el total de la población la distribución fue no normal tanto en la primera prueba como en la segunda.

Discusión

La medición del comportamiento sedentario o la hipoactividad física cobra importancia en la medida que este factor de riesgo influye negativamente sobre los diferentes sistemas del cuerpo humano, y aún más conociendo que el tiempo que se dedica a comportamientos sedentarios en un día está aumentando.

Por otro lado, se sabe que el amplio rango de actividades físicas, desde un comportamiento sedentario como ver televisión en posición sedente hasta la actividad física de intensidad vigorosa, contribuyen al gasto energético total; sin embargo, los investigadores habían estudiado principalmente los efectos benéficos de la actividad física vigorosa y moderada, dejando de lado la actividad física baja e hipoactividad física, y en algunos casos se consideraron como un mismo constructo. Actualmente, se ha cambiado esta visión y las nuevas investigaciones apuntan a encontrar los efectos de la actividad física baja y del comportamiento sedentario sobre la salud de las personas. Para esto, es indispensable contar con instrumentos que reúnan las propiedades psicométricas adecuadas para extraer conclusiones válidas de estas investigaciones.⁴

En este sentido, no puede existir una medición válida a menos que la medición tenga algún grado de confiabilidad, la evaluación de esta propiedad psicométrica es el primer paso para establecer la calidad de las mediciones, de tal modo que la confiabilidad es necesaria pero no suficiente para la validez de una prueba.^{25, 26} Tres términos están involucrados en la confiabilidad: la consistencia interna, la reproducibilidad y el acuerdo.¹⁹ La consistencia interna representa el promedio de las correlaciones entre todos los ítems de una escala,²⁷ la reproducibilidad determina si se obtienen los mismos datos al repetirse la prueba bajo las mismas condiciones¹⁹ y el acuerdo, indica cuán diferentes son los datos entre dos mediciones, dos evaluadores o dos instrumentos en las mismas unidades de medida.¹⁹

Teniendo en cuenta lo anterior, la evaluación de la confiabilidad del tiempo usado en permanecer en posición sedente indagado por el IPAQ-corto aún no ha sido evaluada; puesto que, después de la revisión de la literatura, se encontraron 6 estudios que evaluaron la reproducibilidad, ninguno evaluó el acuerdo de las mediciones de éste ítem en población adulta;⁷⁻¹² y en dos de ellos se evaluó la reproducibilidad por medio del coeficiente de correlación de Spearman,^{7, 8} el cual es una herramienta limitada porque solo evalúa la asociación entre las dos mediciones cuando los datos no se distribuyen de manera normal;¹⁶ mientras que la prueba estadística usada en el presente trabajo (CCI) si evalúa la reproducibilidad.

Reproducibilidad prueba re prueba y nivel de acuerdo del tiempo en posición sedente evaluado mediante el IPAQ y el GPAQ. En el presente estudio la reproducibilidad prueba re prueba de los ítems de hipoactividad del IPAQ-corto medidos en h/d fue de 0.77 (IC95% 0.67-0.84) y no hubo diferencias estadísticamente significativas por género, los resultados son comparables con los estudios de Kurtze *et al*¹¹ y Papatthaniou *et al*¹² con CCI de 0.66 y 0.80, respectivamente. No obstante, el análisis de Bland & Altman muestra un pobre nivel de acuerdo teniendo en cuenta que ver más de 4 horas de televisión está asociado con un aumento de dos veces la probabilidad de presentar sobrepeso comparado con aquellos que ven televisión por una hora, según el estudio de Salmon *et al*²⁸ realizado en adultos de Australia. El promedio de las diferencias fue de -0.04 y los límites de acuerdo de -4.9 y 4.9 h/d, lo cual significa que en promedio en la primera prueba se registró un menor tiempo en posición sedente que en la segunda.

En este estudio el acuerdo compara el tiempo en posición sedente en un día de la semana previa con un intervalo de tiempo de recuerdo entre 3 y 6 días. Los resultados muestran que el tiempo que permanece sentado una persona es diferente en los dos momentos evaluados, lo cual puede deberse a que el patrón de comportamiento haya cambiado o a que el instrumento no sea confiable; por esta razón, vale la pena analizar si realmente con trabajos como este se mide la confiabilidad del ítem de hipoactividad del IPAQ o la consistencia en un comportamiento humano como lo es permanecer sentado. Por otro lado, la variabilidad del tiempo en posición sedente medido mediante el IPAQ corto, que correspondió en la primera prueba a un rango intercuartílico entre 2.4 y 6 h/d y en la segunda prueba a 2.0 y 7.0 h/d, pudo afectar los resultados del CCI, puesto que se sabe que a mayor variabilidad de los datos el CCI se eleva,¹⁸ lo cual no sucede con los límites de acuerdo, de este modo se podrían explicar, en parte, la buena reproducibilidad pero el pobre nivel de acuerdo entre mediciones.

En relación con el GPAQ, solo existe en la literatura un artículo, de reciente publicación, que evalúa su reproducibilidad y validez,³ no obstante, la reproducibilidad del ítem de hipoactividad no fue evaluada. Por lo cual, este es el primer estudio que evalúa la reproducibilidad prueba-re prueba del tiempo en posición sedente medida con este instrumento. Los resultados muestran coeficientes levemente más altos en el GPAQ que en el IPAQ-corto, sin ser estadísticamente diferentes (CCI 0.83, IC95% 0.76-0.89) y un pobre nivel de acuerdo; puesto que, aunque el promedio de las diferencias fue de 0.15 h/d, cercano al cero; los límites de acuerdo fueron amplios (-4.2 y 4.5 h/d). En el GPAQ el acuerdo compara el tiempo en posición sedente en un día de una semana usual, indagado en dos ocasiones con un intervalo de tiempo de recuerdo entre 3 y 6 días, el pobre acuerdo puede explicarse por la dificultad en el recuerdo de una semana usual y de este modo en las dos ocasiones se registre información de un día de dos semanas en las cuales su comportamiento fue diferente.

Reproducibilidad y nivel de acuerdo entre ítems de los cuestionarios. Este es el primer estudio que compara la reproducibilidad del ítem de hipoactividad del IPAQ corto y el GPAQ en población adulta, por lo cual no es posible hacer comparaciones. Los resultados muestran una muy buena reproducibilidad entre cuestionarios tanto para la primera (CCI 0.81, IC95% 0.73-0.87) como para la segunda prueba (CCI 0.82, IC95% 0.74-0.88); no obstante, el nivel de acuerdo es pobre con un promedio de las diferencias de -0.54 h/d (límites de acuerdo: -4.7 a 3.6 h/d) para la primera prueba y de -0.36 h/d (límites de acuerdo -4.9 a 4.2 h/d) para la segunda prueba. Mostrando que en promedio el IPAQ-corto subestima las horas en posición sedente comparadas con las registradas en el GPAQ. Esta tendencia, se puede explicar, en parte, por la diferencia en el tiempo de recuerdo, puesto que el IPAQ-corto indaga por los últimos 7 días mientras que el GPAQ indaga por una semana usual, este último puede mejorar la calidad del recuerdo al indagar por patrones de comportamiento ya adquiridos. Otro aspecto a tener en cuenta es la formulación de la pregunta; mientras el GPAQ indaga por el tiempo pasado viajando en autobús, el IPAQ-corto no lo hace explícitamente, lo cual podría llevar a que se recuerde una actividad más en el GPAQ que aumente el tiempo en posición sedente en este cuestionario.

En síntesis, según los resultados de este estudio, los ítems que miden el tiempo en posición sedente en el IPAQ-corto y el GPAQ tienen una buena reproducibilidad prueba re prueba pero un pobre nivel de acuerdo. Esto último, podría explicarse porque en las dos pruebas puede que no se esté evaluando un patrón de comportamiento similar puesto que, cabe la posibilidad, que el participante recuerde días diferentes en cada una de las pruebas, de este modo podría decirse que aún no se está midiendo el comportamiento sedentario o la hipoactividad física sino que se está llegando a una aproximación de la estimación de ese comportamiento. Por otro lado, los resultados de la reproducibilidad entre cuestionarios muestran que para medir el comportamiento sedentario o hipoactividad física en estudios poblacionales se podría usar cualquiera de las dos preguntas de los cuestionarios; no obstante, el nivel de acuerdo indica que el IPAQ subestima su medición comparado con el GPAQ.

En cuanto a las limitaciones de este estudio, dado que el concepto de reproducibilidad incluye que las dos pruebas se realicen bajo las mismas condiciones, es posible que esta situación no se cumpliera puesto que la primera prueba se realizó en las instalaciones de la universidad y estuvo acompañada de aproximadamente 144 preguntas adicionales, la toma de muestras sanguíneas y mediciones antropométricas; mientras que la segunda prueba se realizó en la residencia del participante y solo se diligenciaron los cuestionarios de actividad física. A su vez, en la primera prueba los últimos cuestionarios en ser diligenciados fueron los de actividad física por lo cual los resultados de la primera prueba se pudieron ver influenciados por el cansancio en el momento de responder y de esta manera disminuir la calidad del recuerdo, subestimando el comportamiento sedentario.

Por otro lado, el promedio de las diferencias de las mediciones no se distribuyó de manera normal en algunos casos; en los cuales, pierde validez el cálculo de los límites de acuerdo al 95% y deben interpretarse cuidadosamente los aquí calculados.¹⁷ Igualmente, se debe tener cuidado al extrapolar los hallazgos del presente trabajo a adolescentes y adultos mayores de 70 años, puesto que se desarrolló en personas entre 18 y 70 años de edad.

Para mejorar la medición del comportamiento sedentario o hipoactividad física se debería aumentar el número de ítems que indaguen por otros comportamientos no incluidos en estos ítems y de este modo mejorar la calidad del recuerdo.

Dado que estudios previos no diferenciaron la actividad física baja del comportamiento sedentario es urgente comenzar a mejorar la medición de éste último para establecer las asociaciones entre el comportamiento sedentario o hipoactividad física y los diferentes desenlaces en salud.

Agradecimientos

A la Universidad Industrial de Santander, específicamente al Departamento de Salud Pública, por el apoyo para el desarrollo de la tesis para obtener el título de Magíster en Epidemiología de la primera autora de este artículo.

Las autoras expresan no tener ningún tipo de conflicto de interés.

Referencias

1. US Department of Health and Human Services. Physical activity and health: A report of the Surgeon General. Atlanta, GA: U.S.: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996.
2. Dietz WH. The role of lifestyle in health: the epidemiology and consequences of inactivity. *Proc Nutr Soc* 1996; 55:829-40.
3. Rosenberg DE, Bull FC, Marshall AL, Sallis JF, Bauman AE. Assessment of sedentary behavior with the International Physical Activity Questionnaire. *J Phys Act Health* 2008; 5:S30-S44.
4. Pate RR, O'Neill JR, Lobelo F. The evolving definition of "sedentary". *Exerc Sport Sci Rev* 2008; 36:173-8.
5. Biddle SJ. Sedentary behavior. *Am J Prev Med* 2007; 33:502-4.
6. Clark BK, Sugiyama T, Healy GN, Salmon J, Dunstan DW, Owen N. Validity and reliability of measures of television viewing time and other non-occupational sedentary behaviour of adults: a review. *Obes Rev* 2009; 10:7-16.
7. Craig CL, Marshall AL, Sjoström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003; 35:1381-95.
8. Bull FC, Tahlia SM, Armstrong T. Global Physical Activity Questionnaire (GPAQ): Nine country reliability and validity study. *J Phys Act Health* 2009; 6:790-804.

9. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). *J Sci Med Sport* 2007; 10:45-51.
10. Deng HB, Macfarlane DJ, Thomas GN, Lao XQ, Jiang CQ, Cheng KK, et al. Reliability and validity of the IPAQ-Chinese: the Guangzhou Biobank Cohort study. *Med Sci Sports Exerc* 2008; 40:303-7.
11. Kurtze N, Rangul V, Hustvedt BE. Reliability and validity of the international physical activity questionnaire in the Nord-Trondelag health study (HUNT) population of men. *BMC Med Res Methodol* 2008; 8:63-71.
12. Papathanasiou G, Georgoudis G, Papandreou M, Spyropoulos P, Georgakopoulos D, Kalfakakou V, et al. Reliability measures of the Short International Physical Activity Questionnaire (IPAQ) in Greek young adults. *Hellenic J Cardiol* 2009; 50:283-94.
13. Ainsworth BE, Bauman A, Benaziza H, Blair SN, Booth ML, Craig CL, et al. USA Spanish version translate 2003, 3-long last 7 days self-administered (version of the IPAQ) Revised October 2002. International Physical Activity Questionnaire 2007. Disponible en: URL: <http://www.ipaq.ki.se/> (fecha de consulta: 10 de agosto, 2007)
14. World Health Organization, Department of Chronic Diseases and Health Promotion. Global Physical Activity Questionnaire (GPAQ). Revised October 2004. World Health Organization 2007. Disponible en: <http://www.who.int/chp/steps/GPAQ/en/index.html> (fecha de consulta: 10 de agosto, 2007)
15. Norman GR, Streiner DL. *Bioestadística*. Madrid: Mosby/Doyma Libros, 1996.
16. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307-10.
18. Szklo M, Nieto FJ. Quality assurance and control. In: Szklo M, Nieto FJ (ed). *Epidemiology beyond the basics*. Gaithersburg: Aspen Publication, 2000: 343-401.
19. Orozco LC. Medición en salud, diagnóstico, evaluación de resultados. Un manual crítico más allá de lo básico. Bucaramanga: Publicaciones UIS, 2009.
20. Keating J, Matyas T. Unreliable inferences from reliable measurements. *Aust J Physiother* 1998; 44:5-10.
21. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
22. CDC/WHO. Epi Info 6.04d 2001. [Software]. Disponible en: <http://www.cdc.gov/epiinfo/Epi6/EI6dwni.htm> (Fecha de acceso: 17 de noviembre de 2007).
23. StataCorp. *Stata statistical software: release 9.1* College Station, 2005.
24. República de Colombia. Ministerio de Salud. Resolución 0008430 de 1993.
25. Latour J, Abaira V, Cabello JB, Lopez SJ. Investigation methods in clinical cardiology. IV. Clinical measurements in cardiology: validity and errors of measurements. *Rev Esp Cardiol* 1997; 50:117-28.
26. Kraemer HC, Periyakoil VS, Noda A. Agreement statistics. Kappa coefficients in medical research. In: D'Agostino RB (ed). *Tutorials in biostatistics. Volume 1: Statistical methods in clinical studies*. John Wiley & Sons: Queensland, 2004:85-105.
27. Streiner DL, Norman GR. *Basic concepts. Health measurement scales. A practical guide to their development and use*. New York: Oxford University Press, 2 ed, 1995:4-14.
26. Salmon J, Bauman A, Crawford D, Timperio A, Owen N. The association between television viewing and overweight among Australian adults participating in varying levels of leisure-time physical activity. *Int J Obes Relat Metab Disord* 2000; 24:600-6.