

ARTÍCULO ORIGINAL

## Validez interna y reproducibilidad de la prueba CAGE en Bucaramanga, Colombia

Oscar F. Herrán<sup>1,2</sup>, María F. Ardila<sup>2</sup>

<sup>1</sup> Escuela de Nutrición y Dietética, Universidad Industrial de Santander, Bucaramanga, Colombia.

<sup>2</sup> Observatorio Epidemiológico de Enfermedades Cardiovasculares, Centro de Investigaciones Epidemiológicas, Universidad Industrial de Santander, Bucaramanga, Colombia.

**Objetivo.** Evaluar la validez interna y la reproducibilidad de *cutting down, annoyance by criticism, guilty feeling, and eye-openers* (CAGE) en Bucaramanga, Colombia, de junio de 2002 a noviembre de 2003.

**Método.** Estudio de consistencia interna y reproducibilidad de tecnología diagnóstica. Ciento nueve sujetos entre 18 y 60 años respondieron en dos ocasiones a CAGE y a una encuesta sobre variables biológicas, sociodemográficas y de dieta. La consistencia interna de CAGE se evaluó con el alfa de Cronbach y la reproducibilidad con el coeficiente de Spearman y el valor kappa de Cohen. Con regresión binomial obtuvimos medidas de asociación entre el riesgo de alcoholismo y las variables biológicas, sociodemográficas y de dieta.

**Resultados.** El riesgo de alcoholismo para la población estudiada fue de 15,6% (IC, 9,0, 22,5). Los hombres presentaron un riesgo aumentado de alcoholismo tres veces mayor que las mujeres ( $p=0,02$ ). El coeficiente alfa de Cronbach fue bajo para todas las preguntas de CAGE (mínimo, 0,08; máximo, 0,28). En la primera aplicación de CAGE se obtuvo un alfa de Cronbach de 0,20, y en la segunda, de 0,27 ( $p=0,226$ ). El coeficiente de correlación de Spearman fue de 0,65 (IC, 0,53, 0,75) para la población; de 0,58 (IC, 0,36, 0,73) para hombres, y de 0,80 (IC, 0,68, 0,88) para mujeres. Los valores de kappas alcanzados están entre moderados y buenos. Ninguna variable biológica, sociodemográfica o de dieta se asoció con el riesgo de alcoholismo.

**Conclusión.** La baja consistencia interna de CAGE es debida a la multidimensionalidad de la prueba; dada su reproducibilidad y estabilidad de las medidas de asociación derivadas, es útil en la investigación epidemiológica en esta población.

**Palabras clave:** alcoholismo/epidemiología, cribado, validez, reproducibilidad de resultados, Colombia.

### Internal validity and reliability of the CAGE test in Bucaramanga, Colombia

**Objective.** The internal validity and the reliability of the CAGE test (*cutting down, annoyance by criticism, guilty feeling, and eye-openers*) was examined for residents of Bucaramanga (northeastern Colombia).

**Methods.** One hundred nine subjects between 18 and 60 years were given the CAGE test on two occasions. Each subject was surveyed for a suite of biological, socioeconomic and dietary variables. The internal consistency of the test was evaluated with Cronbach's alpha, and the reliability with Spearman's correlation coefficient,  $r$ , and Cohen's kappa. A binomial regression produced association measures among the risk of alcoholism and the biological, socioeconomic and dietary variables.

**Results.** The risk of alcoholism in the subjects was of 15.6% (CI 9.0 - 22.5%). The men presented a risk of alcoholism three times that of women ( $p=0.02$ ). Cronbach's alpha was low for all test questions (min=0.08; max=0.28). In the first application of CAGE, the average for Cronbach's alpha was 0.20, in the second application 0.27. The overall Spearman's  $r$  was of 0.65 (CI 0.53 - 0.75), but with significant sexual difference: 0.58 (CI 0.36 - 0.73) for men and 0.80 (CI 0.68 - 0.88) for women. Cohen's kappas ranged from moderate to good. Any biological, socioeconomic and dietary variables was associated positively with the risk of alcoholism.

**Conclusion.** The low internal consistency of CAGE is due to the multiple aspects of the test;

however, because of the high reliability and stability of the association measures, the test was recommended for continued use in epidemiological investigations.

**Key words:** alcoholism/ epidemiology, screening, validity, reproducibility of results.

Una de las herramientas más ampliamente utilizadas en los estudios epidemiológicos es la prueba de *Cutting down, annoyance by criticism, guilty feeling, and eye-openers*, CAGE (1).

La prueba CAGE se derivó de un estudio clínico conducido en Carolina del Norte en 1968 para detectar sujetos alcohólicos mediante cuatro preguntas clave que tienen que ver con la crítica social, la culpabilidad individual, la ingestión matutina y la necesidad sentida de consumir alcohol. Tiene alta especificidad (98%) (2), superior, incluso, a pruebas como la gamma-glutamil transpeptidasa, considerada como la mejor prueba de laboratorio para detectar el exceso del consumo de alcohol (más de 16 tragos por día) o para diagnosticar sujetos alcohólicos (1).

Es una prueba discriminante (para tamizar o clasificar) que se aplica a sujetos consumidores de alcohol, compuesta por cuatro preguntas que toman el valor de 0 cuando la respuesta es negativa, o de 1 si es afirmativa. Dos o tres respuestas afirmativas permiten tener un alto grado de sospecha de alcoholismo; cuatro respuestas positivas son indicio de alcoholismo patológico (1).

La prueba CAGE no sólo es un instrumento confiable, fácil y barato de aplicar, sino también la mejor prueba y la menos intimidante para la clasificación de sujetos alcohólicos (1,3). Sin embargo, a pesar de su amplio uso, su validez es aún controvertida (2,4,5).

La utilidad de cualquier instrumento está determinada por tres elementos complementarios: 1) el grado de relación entre los ítems de la prueba o la evaluación de si alguno puede ser redundante, más conocida como la validez interna,

homogeneidad o consistencia interna; 2) su validez externa o capacidad para clasificar sujetos tan bien o mejor que un referente externo considerado como el estándar de oro, y 3) la capacidad de reproducir en conjunto los resultados en diferentes aplicaciones, o reproducibilidad (6).

Como es difícil sustentar que el efecto del alcohol o la expresión fisiológica de su uso o abuso es diferente en sujetos humanos dada su cultura o su área geográfica, la crítica hacia CAGE no se ha centrado sobre su validez externa. La controversia con CAGE surge de su consistencia interna y de la reproducibilidad o estabilidad de sus resultados en el tiempo.

Los resultados sobre consistencia interna y reproducibilidad de CAGE, como los de cualquier otro instrumento, no tienen un comportamiento estático; dependen del contexto, es decir, de la prevalencia del evento que se va a medir, las expectativas sobre el consumo de bebidas alcohólicas, el grado de entrenamiento de los encuestadores, las circunstancias que correspondan con la pregunta que se va a evaluar o que no lo hagan, las características clínicas y sociodemográficas de los sujetos, el diseño del estudio y de las encuestas complementarias y el grado de enmascaramiento de los evaluadores, etc.

Como se ha demostrado, es necesario evaluar el comportamiento de cualquier prueba cada vez que se utilice en una población diferente de aquella en que se desarrolló (7). A pesar de que lo anterior es bastante lógico, una investigación sobre los estudios de validez en las pruebas de tamizaje y diagnósticas publicados en cuatro importantes revistas, mostró cómo la calidad de la metodología utilizada entre 1978 y 1993 estuvo entre mediocre y buena (8). Esto originó una iniciativa para unificar los criterios sobre cómo reportar los estudios de validez de diagnósticos (9).

La prueba de CAGE cobra relevancia en la medida en que el uso y el abuso del alcohol se viene incrementando en el mundo y, con ello, la morbilidad, la mortalidad, el ausentismo laboral y

**Correspondencia:**

Oscar Fernando Herrán, Calle 33 No. 29-14, apartamento 501, apartado aéreo 2604, Bucaramanga, Colombia.  
Teléfono: (57) 763 20715; telefax: (57) 763 45781  
herran28@intercable.net.co y herran@uis.edu.co

Recibido: 03/06/04; aceptado: 03/02/05

la carga de los servicios de salud por sucesos relacionados (10). Además, porque cada vez es más evidente que la cuantificación del problema del alcoholismo es el primer paso en la comprensión de su compleja red causal.

En Centro y Suramérica, la prueba CAGE es ampliamente utilizada lo cual ha permitido comparaciones del riesgo de alcoholismo en la región (11). En Colombia, la prueba CAGE se ha utilizado desde 1987 como prueba tamiz en las encuestas nacionales sobre alcoholismo y consumo de sustancias psicoactivas (1988, 1993 y 1997) (12,13).

Si partimos de la premisa de que la prueba CAGE tiene validez externa (sensibilidad, 98%) (1,2) y de que el contexto en que se aplique define su utilidad en la investigación epidemiológica, se realizó este estudio con los objetivos de evaluar: 1) la consistencia interna entre los diferentes ítems de CAGE; 2) la reproducibilidad entre dos aplicaciones de CAGE, y 3) la estabilidad de las medidas de asociación alcanzadas entre el riesgo de alcoholismo y variables biológicas, socio-demográficas y de dieta en la población adulta y consumidora de alcohol de una ciudad intermedia de Colombia, Suramérica.

### **Materiales y métodos**

Durante 2002-2003 se realizó un estudio de consistencia interna y reproducibilidad de tecnología diagnóstica en el que, además, se estableció la estabilidad en la asociación entre algunas variables biológicas, sociodemográficas y de dieta con el riesgo de alcoholismo según la prueba CAGE.

**Selección de sujetos.** Por medio de muestreo aleatorio realizado en múltiples etapas, se seleccionaron 109 consumidores de alcohol como representativos por sexo de la población de Bucaramanga. Este tamaño de muestra, calculado con un valor alfa de 0,05 y un poder de 0,90, y asumiendo 10% de pérdidas entre las dos aplicaciones de la prueba, permite obtener coeficientes de correlación entre los puntajes por ítem y total de la prueba CAGE de 0,40 como mínimo (Gilman J, Myatt M. EpiCalc 2000. versión 1.0. Atlanta: Centers for disease Control; 1997).

Este tamaño de muestra también permite calcular hasta el 7% de riesgo alto de alcoholismo si se considera un alfa de 0,01 y un poder de 0,90. Por lo tanto, tuvimos un tamaño de muestra suficiente para dos de las principales medidas de salida del estudio: los coeficientes de correlación y el riesgo de alcoholismo.

Se declararon elegibles los sujetos que cumplieron los siguientes criterios: 1) no se declararon abstemios; 2) reconocieron ser consumidores de alcohol (aquellos que, por lo menos, al mes consumen como mínimo un trago, 1 onza o 30 ml de cualquier licor o una botella de cerveza); 3) alfabetos, y 4) que hubieran residido dos años en la ciudad, como mínimo.

De un mapa de Bucaramanga se eligieron al azar manzanas y, luego, casas por estrato socio-económico; en cada vivienda se realizó un censo de los sujetos elegibles y se seleccionó aleatoriamente uno por vivienda. Se encuestaron después de haberlos seleccionado e informado de los objetivos y de haber obtenido su consentimiento escrito.

**Fuentes de información.** Se aplicaron dos encuestas por encuestadores previamente entrenados en la recolección de datos en estudios de población: 1) una encuesta sobre factores biológicos, incluido el índice de masa corporal (IMC) (14), sociodemográficos y de dieta, incluida la frecuencia del consumo de bebidas alcohólicas y el gusto por ellas; 2) prueba de CAGE, ampliamente utilizada en estudios de población y clínicos (1,15). Esta última se aplicó en el hogar y en dos ocasiones al mismo sujeto con un intervalo de dos meses, con el fin de que las respuestas de la segunda aplicación no estuvieran influenciadas por la memoria.

**Calidad de los datos.** Los cuatro ítems de la prueba CAGE se retomaron de los utilizados en las encuestas nacionales sobre consumo. Durante la primera visita, y de manera secuencial, se realizó la encuesta sobre variables biológicas, socioeconómicas y de dieta; por último, se aplicó la prueba CAGE. El gusto manifestado por las bebidas alcohólicas fue el reporte sobre la percepción que del consumo tiene cada sujeto, el cual corresponde a una representación social (16).

Para permitir posteriores comparaciones de los resultados, el riesgo de alcoholismo se definió como "bajo" cuando se obtuvo un puntaje inferior de 3 en la prueba de CAGE, y de "alto" cuando obtuvo un puntaje de 3 o 4; este punto de corte es el propuesto en la versión original de CAGE (1) y el utilizado en el país en las encuestas nacionales sobre consumo (12,13).

Ni los sujetos encuestados ni los encuestadores conocieron los resultados de CAGE y los datos sólo se conocieron por los códigos en el momento de hacerse el análisis estadístico.

**Análisis estadístico.** La normalidad de las variables continuas se evaluó utilizando las pruebas de Shapiro-Wilk y Shapiro-Francia (17). Para describir las categorías de las variables biológicas, socioeconómicas y de dieta se calcularon medidas descriptivas y sus intervalos de confianza al 95% (IC95%).

Para establecer si existían diferencias estadísticamente significativas entre las diferentes categorías por sexo o aplicaciones de CAGE, se calcularon las pruebas t de Student, las pruebas de Anova y de  $\chi^2$ . Para determinar la consistencia interna de CAGE, además del cálculo del coeficiente alfa de Cronbach, se estimaron los coeficientes de correlación entre los ítem, los coeficientes de correlación para cada ítem versus los restantes (18) y las medidas descriptivas para cada pregunta de CAGE y el puntaje acumulado.

La reproducibilidad (prueba-reprueba) entre las dos aplicaciones de CAGE se estimó por medio de: 1) el cálculo de coeficientes de correlación de Spearman (19) y sus respectivos IC95%, entre los puntajes alcanzados para cada ítem y el total acumulado; 2) la estabilidad en las medidas de asociación entre el riesgo de alcoholismo por CAGE y las variables biológicas, socioeconómicas y de dieta (razones de prevalencia) obtenidas mediante regresión binomial, y 3) el cálculo de coeficientes según la kappa de Cohen, que son equivalentes a los coeficientes de correlación intraclase (20).

Para este coeficiente se estimó su máximo valor posible (valor máximo de kappa) dado el acuerdo observado (Po), logrado al reemplazar en el

denominador de la fórmula clásica de Cohen, el valor de Po en lugar de 1:

valor máximo de kappa =  $(Po - Pe) / \text{valor máximo de } (Po - Pe)$ ,

donde Po es el acuerdo observado y Pe es el acuerdo esperado (21). Además, los coeficientes de kappa de Cohen se ajustaron por el sesgo entre observadores (22) que se obtiene al reemplazar en una tabla de contingencia los valores de b y c por su media  $m = (b+c)/2$  para, luego, calcular el coeficiente kappa:

|              |    | Observador A |    |
|--------------|----|--------------|----|
|              |    | Sí           | No |
| Observador B | Sí | a            | b  |
|              | No | c            | d  |

El ajuste del coeficiente kappa también se calculó con la prevalencia alcanzada de alto riesgo de alcoholismo en cada una de las dos mediciones (PABAK) (22) que se obtiene al reemplazar, además de lo anterior, los valores de a y d por su media  $m_1 = (a+d)/2$ :

$$PABAK = [(2m_1/N) - 0,5 / (1 - 0,5)].$$

En términos de Eggin *et al.*, los índices de concordancia de kappa se ajustaron por las características del contexto (23).

Todos los registros se digitaron por duplicado y se limpiaron con la aplicación Validate de EpiInfo 6.04d (Centers for Disease Control. EpiInfo, versión 6.04d. Epidemiología en ordenadores. Atlanta, GA: CDC; 2001). El procesamiento de datos y su análisis se hizo con Stata 8.2/SE (StataCorp. Stata Statistical Software, release 8.2. College Station, TX: Stata Corporation; 2003).

El Comité de Ética de la Facultad de Salud aprobó el protocolo de investigación.

## Resultados

Dado que los valores estimados no se modificaron al corregir por el efecto del diseño de la muestra (24), los reportados son similares a los obtenidos en un muestreo aleatorio simple.

Los 109 sujetos respondieron la prueba CAGE en dos ocasiones. El tiempo medio entre cada aplicación fue de 56 días (IC95%, 55,6, 56,4). La

**Cuadro 1.** Características de la población estudiada. Validez interna y reproducibilidad de CAGE en Bucaramanga. Colombia, 2003.

| Variable                     | Total              | Hombre [n=54]     | Mujer [n=55]      | Valor p |
|------------------------------|--------------------|-------------------|-------------------|---------|
| Edad (años)                  | 27,8 (25,8, 29,8)* | 29,8 (26,7, 32,8) | 25,8 (23,2, 28,5) | 0,06    |
| Escolaridad (años)           | 10,4 (9,7, 11,1)   | 10,3 (9,2, 11,4)  | 10,5 (9,6, 11,4)  | 0,82    |
| Peso (kg)                    | 61,8 (59,6, 60,0)  | 67,8 (64,7, 70,9) | 55,9 (53,6, 58,2) | 0,00    |
| IMC (kg/talla <sup>2</sup> ) | 22,6 (21,9, 23,2)  | 23,1 (22,0, 24,0) | 22,0 (21,1, 22,9) | 0,10    |
| > a 25                       | 25 (22,9)†         | 14 (25,9)         | 11 (20,0)         | 0,31    |
| Diagnóstico médico actual‡   | 13 (11,9)          | 4 (7,4)           | 9 (16,4)          | 0,15    |
| Hizo dieta¶                  | 11 (10,0)          | 5 (9,3)           | 6 (10,9)          | 0,77    |
| Gusto por el trago           |                    |                   |                   | 0,96    |
| Poco                         | 39 (35,8)          | 20 (37,0)         | 19 (34,6)         |         |
| Moderado                     | 64 (58,7)          | 31 (57,4)         | 33 (60,0)         |         |
| Mucho                        | 6 (5,5)            | 3 (5,6)           | 3 (5,4)           |         |
| Frecuencia de consumo        |                    |                   |                   | 0,07    |
| Semanal                      | 24 (22,9)          | 17 (28,5)         | 8 (14,6)          |         |
| Quincenal                    | 27 (24,8)          | 16 (29,6)         | 11 (20,0)         |         |
| Mensual                      | 57 (52,3)          | 21 (41,9)         | 36 (65,4)         |         |
| Riesgo de alcoholismo        | 17 (15,6)          | 13 (24,1)         | 4 (7,3)           | 0,02    |

\* Promedio e IC95%. † Frecuencia (porcentaje). ‡ Enfermedad o trastorno con diagnóstico hecho por un médico.

¶ En el último año. | CAGE >= a 3.

**Cuadro 2.** Consistencia interna entre ítems de CAGE.

| Ítem†          | Primera aplicación* |             |                        |                          | Segunda aplicación* |             |             |                        |                          |                   |
|----------------|---------------------|-------------|------------------------|--------------------------|---------------------|-------------|-------------|------------------------|--------------------------|-------------------|
|                | Media               | DE          | Correlación ítem-rest‡ | Correlación inter-ítem § | Alfa¶<br>Cronbach   | Media       | DE          | Correlación ítem-rest‡ | Correlación inter-ítem § | Alfa¶<br>Cronbach |
| Pregunta 1     | 0,25                | 0,43        | 0,106                  | 0,053                    | 0,144               | 0,33        | 0,47        | 0,209                  | 0,046                    | 0,127             |
| Pregunta 2     | 0,77                | 0,42        | 0,153                  | 0,027                    | 0,077               | 0,43        | 0,50        | 0,174                  | 0,066                    | 0,174             |
| Pregunta 3     | 0,53                | 0,50        | 0,052                  | 0,085                    | 0,219               | 0,71        | 0,45        | 0,090                  | 0,115                    | 0,280             |
| Pregunta 4     | 0,61                | 0,49        | 0,077                  | 0,070                    | 0,184               | 0,59        | 0,49        | 0,085                  | 0,118                    | 0,285             |
| <b>Todas**</b> | <b>1,84</b>         | <b>0,77</b> |                        | <b>0,059</b>             | <b>0,200</b>        | <b>1,93</b> | <b>0,66</b> |                        | <b>0,086</b>             | <b>0,274</b>      |

\* Para todas las preguntas n=109. † Para todas: mínimo=0 (no), máximo=1 (sí). ‡ Correlación entre el ítem y el resto de ítems. § Promedio de la correlación inter-ítem. | *t de Student* entre aplicaciones ( $p < 0,001$ ). ¶ Alfa de Cronbach, si se eliminara esa pregunta de la prueba. \*\*Valores medios para la prueba de CAGE; *t de Student* para medias globales alcanzadas entre aplicaciones ( $p = 0,337$ ); *t de Student* para alfa de Cronbach alcanzadas entre aplicaciones ( $p = 0,226$ ). **Pregunta 1:** ¿En el último año ha sentido alguna vez la necesidad de disminuir la cantidad de alcohol que toma? **Pregunta 2:** ¿En el último año se ha sentido molesto porque le critican su manera de beber? **Pregunta 3:** ¿En el último año se ha sentido alguna vez mal o culpable por su manera de beber? **Pregunta 4:** ¿En el último año ha tomado alguna vez un trago en la mañana para calmar los nervios o quitar el guayabo?

edad media de los 54 hombres encuestados fue mayor en 3 años que la de las mujeres ( $p = 0,06$ ). Al igual que la edad, la frecuencia de consumo de bebidas alcohólicas fue mayor en los hombres ( $p = 0,07$ ). El gusto manifestado por las bebidas alcohólicas no es diferente según el sexo ( $p = 0,96$ ). La frecuencia de consumo semanal de alcohol en los hombres fue el doble del de las mujeres (28,5%) y la quincenal, un tercio mayor (29,6%). El alto riesgo de alcoholismo según CAGE para toda la

población fue de 15,6% (IC95%, 9,0, 22,5). Los hombres presentaron un riesgo aumentado de alcoholismo tres veces mayor que las mujeres ( $p = 0,02$ ) (cuadro 1).

**Consistencia interna de CAGE.** En la primera aplicación de la prueba CAGE, la segunda pregunta "¿En el último año se ha sentido molesto porque le critican su manera de beber?" fue la que alcanzó el mayor número de respuestas negativas, 77%. En la segunda aplicación, fue la

tercera pregunta con 72%: "¿En el último año se ha sentido alguna vez mal o culpable por su manera de beber?" La primera pregunta de CAGE; "¿En el último año ha sentido alguna vez la necesidad de disminuir la cantidad de alcohol que toma?" fue la que alcanzó en las dos aplicaciones el menor número de respuestas negativas, 25% y 33%, respectivamente. La diferencia media en el puntaje total alcanzado en cada una de las dos aplicaciones de CAGE fue de -0.09 (IC95%, -0,18, -0,00) ( $p=0,05$ ).

La correlación entre ítems fue baja para todas las preguntas en las dos aplicaciones (mínima: 0,03; máxima: 0,21). El coeficiente alfa de Cronbach fue bajo para todas las preguntas de la prueba CAGE (mínimo: 0,08; máximo: 0,28). En la segunda aplicación se obtuvo un aumento de 27% en el coeficiente alfa de Cronbach para el puntaje total con respecto al alcanzado en la primera aplicación, 0,20 versus 0,27, respectivamente ( $p=0,226$ ). De manera complementaria, los puntajes de la media global entre aplicaciones no fueron diferentes ( $p=0,337$ ) (cuadro 2).

**Reproducibilidad de CAGE.** Al comparar los puntajes totales obtenidos entre aplicaciones se obtuvo un coeficiente de correlación de Spearman ( $r_s$ ) de 0,65 (IC95%, 0,53, 0,75) para toda la población, de 0,58 (IC95%, 0,36, 0,73) para hombres y 0,80 (IC95%, 0,68, 0,88) para mujeres. La cuarta pregunta, "¿En el último año ha tomado alguna vez un trago en la mañana para calmar los nervios o quitar el guayabo?", fue la que presentó el  $r_s$  mayor, de manera consistente para toda la población, los hombres y las mujeres. Para la

segunda y tercera preguntas, los  $r_s$  alcanzados fueron pobres, mientras que para la primera y cuarta preguntas pueden considerarse como aceptables (cuadro 3).

Al evaluar el grado de acuerdo obtenido para el total de los puntajes por CAGE, se observaron coeficientes de kappa de Cohen considerados como buenos y muy buenos (mayores de 0,60). La razón entre el coeficiente kappa de Cohen y el coeficiente kappa máximo fue de 100% para toda la población, 89% para los hombres y 78,1% para las mujeres. El coeficiente kappa corregido por el sesgo entre mediciones y por la prevalencia de alto riesgo de alcoholismo fue de 82% para toda la población, 70% para los hombres y 93% para las mujeres (cuadro 4).

La segunda y la tercera pregunta tuvieron los más bajos porcentajes de respuestas positivas (mínimo: 18,2%; máximo: 50,0%). Al analizar la concordancia de los ítems de la prueba CAGE para estas dos preguntas, observamos que el coeficiente kappa de Cohen, así como la razón entre el kappa y el kappa máximo y la PABAK, alcanzaron consistentemente los menores valores para toda la población, hombres y mujeres (cuadro 4).

**Estabilidad de las medidas de asociación entre las variables biológicas, socioeconómicas y de dieta con la prueba CAGE.** No hubo diferencias entre aplicaciones para ninguna de las proporciones de alto riesgo de alcoholismo obtenidas por CAGE ( $p>0,05$ ). En la primera aplicación, el alto riesgo de alcoholismo según CAGE fue menor en las mujeres, razón de

**Cuadro 3.** Coeficientes de correlación de Spearman ( $r_s$ ) por sexo, entre dos aplicaciones de CAGE.

| Ítems      | Total<br>$r_s$ (IC95%) | Hombre<br>$r_s$ (IC95%) | Mujer<br>$r_s$ (IC95%) |
|------------|------------------------|-------------------------|------------------------|
| Pregunta 1 | 0,46 (0,29, 0,59)†     | 0,47 (0,23, 0,65)†      | 0,44 (0,19, 0,63)†     |
| Pregunta 2 | 0,26 (0,07, 0,42)†     | 0,26 (-0,01, 0,49)      | 0,24 (-0,03, 0,48)     |
| Pregunta 3 | 0,22 (0,04, 0,39)†     | 0,15 (-0,12, 0,40)      | 0,29 (0,03, 0,52)†     |
| Pregunta 4 | 0,54 (0,40, 0,66)†     | 0,56 (0,35, 0,72)†      | 0,48 (0,24, 0,66)†     |
| CAGE       | 0,65 (0,53, 0,75)†     | 0,58 (0,36, 0,73)†      | 0,80 (0,68, 0,88)†     |

†  $p<0,05$ . **Pregunta 1:** ¿En el último año ha sentido alguna vez la necesidad de disminuir la cantidad de alcohol que toma? **Pregunta 2:** ¿En el último año se ha sentido molesto porque le critican su manera de beber? **Pregunta 3:** ¿En el último año se ha sentido alguna vez mal o culpable por su manera de beber? **Pregunta 4:** ¿En el último año ha tomado alguna vez un trago en la mañana para calmar los nervios o quitar el guayabo?

**Cuadro 4.** Acuerdo alcanzado entre dos aplicaciones de CAGE.

| Ítem                         | Sí (%) | Kappa* | Kap. máx.† | Razón‡ | BAK   | PABAK** |
|------------------------------|--------|--------|------------|--------|-------|---------|
| <b>Total</b>                 |        |        |            |        |       |         |
| Pregunta 1                   | 82,8   | 0,266  | 0,362      | 73,6   | 0,257 | 0,495   |
| Pregunta 2                   | 22,9   | 0,197  | 0,536      | 36,8   | 0,101 | 0,138   |
| Pregunta 3                   | 46,8   | 0,207  | 0,333      | 62,3   | 0,179 | 0,229   |
| Pregunta 4                   | 39,4   | 0,543  | 0,564      | 96,2   | 0,543 | 0,560   |
| <b>Hombres</b>               |        |        |            |        |       |         |
| Pregunta 1                   | 72,2   | 0,456  | 0,576      | 79,1   | 0,450 | 0,519   |
| Pregunta 2                   | 27,8   | 0,213  | 0,509      | 41,8   | 0,134 | 0,148   |
| Pregunta 3                   | 50,0   | 0,148  | 0,200      | 74,1   | 0,134 | 0,148   |
| Pregunta 4                   | 50,0   | 0,556  | 0,652      | 85,2   | 0,553 | 0,556   |
| <b>Mujeres</b>               |        |        |            |        |       |         |
| Pregunta 1                   | 78,2   | 0,429  | 0,530      | 81,0   | 0,425 | 0,564   |
| Pregunta 2                   | 18,2   | 0,175  | 0,560      | 31,3   | 0,057 | 0,127   |
| Pregunta 3                   | 43,6   | 0,252  | 0,516      | 48,8   | 0,204 | 0,309   |
| Pregunta 4                   | 21,8   | 0,458  | 0,628      | 72,9   | 0,450 | 0,564   |
| <b>CAGE (3 o más puntos)</b> |        |        |            |        |       |         |
| Total                        | 15,6   | 0,652  | 0,652      | 100    | 0,652 | 0,817   |
| Hombres                      | 24,1   | 0,572  | 0,641      | 89,3   | 0,571 | 0,704   |
| Mujeres                      | 7,3    | 0,781  | 1,00       | 78,1   | 0,78  | 0,927   |

\* Kappa de Cohen (0,5). † Máximo valor posible para la kappa de Cohen. ‡ Razón (%) entre la kappa de Cohen y la kappa máxima. | Kappa de Cohen ajustada por sesgo entre observadores. \*\* Kappa de Cohen ajustada por la prevalencia y el sesgo entre observadores. **Pregunta 1:** ¿En el último año ha sentido alguna vez la necesidad de disminuir la cantidad de alcohol que toma? **Pregunta 2:** ¿En el último año se ha sentido molesto porque le critican su manera de beber? **Pregunta 3:** ¿En el último año se ha sentido alguna vez mal o culpable por su manera de beber? **Pregunta 4:** ¿En el último año ha tomado alguna vez un trago en la mañana para calmar los nervios o quitar el guayabo?

prevalencia de 0,30 (IC95%, 0,10, 0,87) ( $p=0,03$ ), en la segunda, aun cuando se mantuvo la tendencia, no fue estadísticamente significativa, razón de prevalencia de 0,54 (IC95%, 0,21, 1,34). Ninguna otra razón de prevalencia estimada para las variables biológicas, socioeconómicas y de dieta fue diferente entre aplicaciones (cuadro 5).

### Discusión

Debido al diseño y los métodos de recolección utilizados, los resultados se derivaron de la población adulta residente en la zona urbana de Bucaramanga, alfabeta, consumidora de alcohol y de los seis diferentes estratos socioeconómicos. Sin embargo, debido al amplio rango de edad de los encuestados y a que no hay datos confiables sobre esta variable para Bucaramanga, no pudimos verificar si en la muestra sigue la misma distribución que la de la población general. Debido al diseño, el nivel de educación fue probablemente mayor en nuestra muestra. Por lo tanto, la inferencia de los resultados podría estar limitada para los sujetos con niveles bajos de educación,

aun cuando la proporción de sujetos sin terminar la primaria en nuestra muestra (5,5%) es muy similar a la de la población general, 4,7% (25).

El 15,6% de la población estudiada está en alto riesgo de alcoholismo, más del doble de lo reportado para el país, 7,1% (IC95%, 5,5, 7,5) en el Estudio Nacional de Salud Mental y consumo de sustancias psicoactivas de Colombia (12) y 3% más que lo reportado en 1997 (13). En el Estudio Nacional de Salud Mental por cada 6 hombres se encontraba una mujer en alto riesgo. En Bucaramanga, por cada 3 hombres hay una mujer; este aumento del riesgo de alcoholismo en las mujeres también se reportó en 1997 (13).

La prueba CAGE es, tal vez, el instrumento menos intimidante para tamizar consumidores con alto riesgo de alcoholismo y, a pesar de que se enfatiza en que el resultado positivo a CAGE no es un diagnóstico de alcoholismo, tres o cuatro respuestas positivas alcanzan una especificidad y valor pronóstico positivo superiores a 0,99 (26). Otros investigadores han utilizado puntos de corte

**Cuadro 5.** Razones de prevalencia para riesgo de alcoholismo, alcanzadas en dos aplicaciones de CAGE.

| Ítem*                        | Primera aplicación |                   | Segunda aplicación |                   |
|------------------------------|--------------------|-------------------|--------------------|-------------------|
|                              | % CAGE†            | RP (IC95%)        | % CAGE†            | RP (IC95%)        |
| Edad (años)                  |                    |                   |                    |                   |
| 30 o menos                   | 15,8               | 1,0               | 17,1               | 1,0               |
| Más de 30                    | 15,2               | 0,96 (0,37, 2,51) | 12,1               | 0,71 (0,25, 2,01) |
| Sexo                         |                    |                   |                    |                   |
| Hombre                       | 24,1               | 1,0               | 20,4               | 1,0               |
| Mujer                        | 7,3                | 0,30 (0,10, 0,87) | 10,9               | 0,54 (0,21, 1,34) |
| IMC (kg/talla <sup>2</sup> ) |                    |                   |                    |                   |
| 25 o menos                   | 16,5               | 1,0               | 17,6               | 1,0               |
| Más de 25                    | 12,5               | 0,76 (0,24, 2,42) | 8,3                | 0,47 (0,12, 1,92) |
| Escolaridad (años)           |                    |                   |                    |                   |
| 11 o menos                   | 17,1               | 1,0               | 18,6               | 1,0               |
| Más de 11                    | 12,8               | 0,75 (0,28, 1,97) | 10,3               | 0,55 (0,19, 1,58) |
| Frecuencia de consumo        |                    |                   |                    |                   |
| Semanal                      | 20,0               | 1,0               | 20,0               | 1,0               |
| Quincenal                    | 22,2               | 1,11 (0,39, 3,19) | 29,6               | 1,5 (0,56, 3,93)  |
| Mensual                      | 10,5               | 0,53 (0,18, 1,56) | 7,0                | 0,35 (0,10, 1,20) |
| Gusto por el trago           |                    |                   |                    |                   |
| Poco                         | 15,4               | 1,0               | 17,9               | 1,0               |
| Moderado                     | 14,1               | 0,91 (0,35, 2,37) | 10,9               | 0,61 (0,23, 1,61) |
| Mucho                        | 33,3               | 2,2 (0,56, 8,36)  | 50,0               | 2,8 (0,98, 7,91)  |

\* En la primera aplicación se evidenciaron diferencias entre sexo ( $p=0,026$ ); en la segunda aplicación, entre categorías de consumo ( $p=0,022$ ) y gusto por el trago ( $p=0,037$ ). Entre aplicaciones no hubo diferencias para ninguna variable o subcategoría.  
† CAGE  $\geq$  a 3.

diferentes al de este estudio, declarando el riesgo alto de alcoholismo con dos o, incluso, una sola respuesta positiva en la prueba CAGE (27-30).

El punto de corte es una consecuencia ligada al contexto en que se aplique la prueba CAGE, pues se ha demostrado que cuando se hacen encuestas sobre el consumo de bebidas alcohólicas previas a la aplicación de CAGE, o cuando se realizan actividades de desensibilización, con dos respuestas positivas se logra alcanzar una especificidad de hasta el 95% (31).

A pesar de que nosotros aplicamos una encuesta sociodemográfica previa a la prueba CAGE e indagamos por variables biológicas y de dieta, nuestro caso no es el descrito anteriormente y, por lo tanto, un punto de corte de tres o más respuestas positivas es el recomendado (31).

En forma complementaria, dado que el principal uso de la prueba CAGE es como prueba tamiz y a que se ha demostrado que una o dos respuestas positivas, alcanzan el 62% y el 82%, respectivamente del valor pronóstico positivo, frente al 99%

cuando se declara el riesgo alto con tres o cuatro respuestas positivas (2), este último es un punto de corte más sensible. Además, permite comparaciones con las encuestas nacionales (12,13).

En cualquier circunstancia, una o dos respuestas positivas a la prueba CAGE reflejan en buena medida un consumo de alcohol problemático, sin que se alcance la categoría de alcohólico (32).

La consistencia interna fue baja para cada uno de los ítems y en conjunto. Además, no fue estadísticamente diferente entre aplicaciones ( $p=0,226$ ). Una baja consistencia interna es previsible en cuestionarios con pocos ítems como CAGE y podría estar asociada, en este caso particular, a la multidimensionalidad de la prueba (33). Este hallazgo es particularmente sorprendente, pues en otras evaluaciones de la consistencia interna de la prueba CAGE en población colombiana, el coeficiente alfa de Cronbach fue sustancialmente mayor (0,69 a 0,75) (13, 32); estos valores se obtuvieron cuando la prueba

CAGE se aplicó inmerso en otras preguntas sobre consumo, pero los autores aclaran que la confiabilidad es mayor cuando se aplica antes de otras relacionadas con el consumo de bebidas alcohólicas, como se discutió anteriormente.

La aplicación repetida de la prueba CAGE, aun cuando aumentó ligeramente el coeficiente alfa de Cronbach para el conjunto y cada ítem, no cambió su interpretación "clásica" de inaceptable. Sin embargo, si entendemos que una alta consistencia interna se obtiene cuando los ítems de una prueba miden una misma expresión o dimensión lo que se debe esperar lógicamente en la prueba CAGE es una consistencia interna baja, porque la prueba CAGE, además de tener pocos ítems, es multidimensional. La consistencia interna vista de esta manera no es una medida de la calidad de una prueba, sino de su dimensionalidad, elemento que es muy importante en el diseño de la prueba y que permite evaluar su correspondencia con el sustento teórico que la soporta, pero poco relevante cuando se pasa al uso en la investigación aplicada.

Un interés particular de este estudio fue evaluar la reproducibilidad de la prueba CAGE en aplicaciones repetidas con suficiente tiempo entre ellas, para librarlas de la memoria. El coeficiente de correlación de Spearman fue aceptable para la población total y los hombres, y bueno para las mujeres. El alto  $r_s$  en las mujeres está condicionado por su baja proporción de alto riesgo de alcoholismo. La cuarta pregunta que es menos intimidante en términos del autorreconocimiento de la culpabilidad y la que menos juzgamiento moral y social puede derivar por la cultura hacia el consumo de bebidas alcohólicas, fue la que mejor reproducibilidad alcanzó. En el mismo sentido, la primera pregunta que puede ser liberadora del sentimiento de culpa y reivindicante, fue la que mayor proporción de respuestas positivas alcanzó (cuadro 4).

Otra forma de evaluar la reproducibilidad de la prueba CAGE fue el medir el grado de concordancia alcanzado entre aplicaciones mediante los coeficientes kappa. La concordancia para cada ítem, en general, puede considerarse como baja. Sin embargo, es muy buena para la prueba CAGE en conjunto, incluso corregida por

el sesgo entre observador y por la prevalencia o contexto (PABAK). Lo anterior está a favor de una alta estabilidad de los resultados al usar la prueba CAGE en la investigación epidemiológica, elemento deseable para cualquier prueba.

El cálculo de las razones de prevalencia confirmó que los resultados con las variables biológicas, socioeconómicas y de dieta son reproducibles y muy estables cuando se usa la prueba CAGE con un punto de corte de 3 o más (cuadro 5). Bajo la premisa de que la prueba CAGE cuenta con validez externa, su reproducibilidad junto con la estabilidad de las medidas de asociación derivadas de ella, son la mejor expresión de la utilidad en la investigación epidemiológica aplicada.

Con excepción del sexo, en la primera aplicación ninguna de las variables estudiadas se asoció con el riesgo de alcoholismo (cuadro 5). Cada vez es más evidente que las aproximaciones sociodemográficas no permiten comprender ni intervenir el riesgo de alcoholismo y que la mejor alternativa es el estudio de las expectativas, definidas como "la anticipación de una relación sistemática entre eventos u objetos en una situación futura" (34). El estudio de las expectativas ha sido uno de los principales aportes contemporáneos en el estudio de las adicciones (35,36).

El Estudio Nacional de Salud Mental afirma que las variables asociadas a la esfera familiar, las redes de apoyo, la valoración social del consumo de alcohol y de las bebidas alcohólicas, las relaciones intra e interfamiliares, la relación de pareja, la percepción sobre la autoridad y las autoridades, el concepto de lo público y de ciudadanía y otras igualmente complejas de medir, son las determinantes del consumo de alcohol y del riesgo de alcoholismo.

Los diseños y los métodos para el estudio de tecnología diagnóstica están aún en desarrollo (37). Aquí se han presentado algunos índices considerados como los mejores para evaluar el comportamiento de una prueba que es ampliamente usada en población adulta y consumidora de alcohol. Este es el primer paso para el estudio de los condicionantes del riesgo de alcoholismo.

Los resultados de este estudio permiten afirmar que la baja consistencia interna de la prueba de CAGE es debida a la multidimensionalidad del constructo que lo respalda, lo que se convierte en una fortaleza de la prueba; además, dada su alta reproducibilidad y estabilidad de las medidas de asociación con variables biológicas, socioeconómicas y de dieta, la prueba CAGE es muy útil en la investigación epidemiológica que se realice en la población adulta de Bucaramanga.

Podría afirmarse que la prueba CAGE como un todo permite medidas de asociación confiables con variables biológicas, socioeconómicas y de dieta, no así cuando éstas se relacionan con sus ítems particulares. Es importante resaltar una vez más que los resultados aquí obtenidos fueron con un punto de corte de 3 o más, pues como se discutió previamente, algunos reportes usan otro menos conservador de una o dos respuestas positivas. Finalmente, sería deseable el diseño y la ejecución de estudios sobre la validez externa de la prueba con el fin de conocer los valores de sensibilidad, especificidad y los valores pronósticos que alcanza la prueba CAGE en la población colombiana.

#### Conflicto de intereses

Los autores manifestamos que no existe ningún conflicto de interés, ni real ni potencial sobre los resultados presentados.

#### Financiación

Este trabajo fue financiado por el Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología "Francisco José de Caldas" (Colciencias, contrato No. 127-2002; código 1102-04-11720) y la Universidad Industrial de Santander.

#### Referencias

1. **Ewing JA.** Detecting alcoholism. The CAGE questionnaire. *JAMA* 1984;252:1905-7.
2. **Bisson J, Nadeau L, Demers A.** The validity of the CAGE scale to screen for heavy drinking and drinking problems in a general population survey. *Addiction* 1999;94:715-22.
3. **Soderstrom CA, Smith GS, Kufera JA, Dischinger PC, Hebel JR, McDuff DR et al.** The accuracy of the CAGE, the Brief Michigan Alcoholism Screening Test, and the Alcohol Use Disorders Identification Test in screening trauma center patients for alcoholism. *J Trauma* 1997;43:962-9.
4. **Knight JR, Sherritt L, Harris SK, Gates EC, Chang G.** Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. *Alcohol Clin Exp Res* 2003;27:67-73.
5. **Watkins JP, Eisele GR, Matthews KO.** Occupational medical program alcohol screening. Utility of the CAGE and BMAST. *J Subst Abuse Treat* 2000;19:51-7.
6. **Bland JM, Altman DG.** Statistics notes: validating scales and indexes. *BMJ* 2002;324:606.
7. **Chmura H.** Evaluating medical test. London: Sage Publications; 1992. p.266-85.
8. **Reid MC, Lachs MS, Feinstein AR.** Use of methodological standard in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-51.
9. **Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al.** The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.
10. **Klatsky AL, Armstrong AA, Friedman GD.** Alcohol and mortality. *Ann Intern Med* 1992;117:646-54.
11. **Orpinas P.** Who is violent? Factors associated with aggressive behaviors in Latin America and Spain. *Rev Panam Salud Pública* 1999;5:232-44.
12. **Torres Y, Posada V, Rojas MC.** Estudio Nacional de Salud Mental y Consumo de Sustancias Sicoactivas. Bogotá: Ministerio de Salud; 1993.
13. **Torres de GY.** Alcohol: prevalencia de consumo y dependencia en Colombia. *Revista de Medicina CES* 1999;12:1-9.
14. **Bray GA.** Obesity in America. An overview of the Second Fogarty International Conference on obesity. *Int J Obes* 1979;3:363-75.
15. **Tate DG.** Alcohol use among spinal cord-injured patients. *Am J Phys Med Rehabil* 1993;72:192-5.
16. **Vázquez G, Prada GE, Herrán OF.** ¿Cuánto se percibe? Una aproximación a la ingesta dietaria. *Salud UIS* 2002;34:89-98.
17. **Gould WW.** Final summary of test of normality. *Stata Technical Bulletin* 1992;5:10-1.
18. **Reynaldo J, Santos A.** Cronbach's alpha: tool for assessing the reliability of scales. *Journal of Extensión* 1999; 37. Disponible en: <http://joe.org/joe/1999april/tt3.html>. Acceso el 10 de mayo de 2004.
19. **Altman DG.** Practical statistics for medical research. New York: Chapman & Hall/CRC; 1999. p.285-8.
20. **Bartfay E, Donner A.** The effect of collapsing multinomial data when assessing agreement. *Int J Epidemiol* 2000; 29:1070-5.

21. **Feinstein AR, Cicchetti DV.** High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
22. **Byrt T, Bishop J, Carlin JB.** Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423-9.
23. **Eggin TK, Feinstein AR.** Context bias. A problem in diagnostic radiology. *JAMA* 1996;276:1752-5.
24. **Eltinge JL, Sribney WM.** Estimation of means, totals, ratios, and proportions for survey data. *Stata Technical Bulletins Reprints* 1996;6:213-35.
25. **Profamilia.** Encuesta Nacional de Demografía y Salud. Resumen Región Oriental, 2000. Bogotá, Colombia: Profamilia; 2000.
26. **Bush B, Shaw S, Cleary P, Delbanco TL, Aronson M.** Screening for alcohol abuse using the CAGE questionnaire. *Am J Med* 1987;82:231-5.
27. **Colhoun H, Ben-Shlomo Y, Dong W, Bost L, Marmot M.** Ecological analysis of collectivity of alcohol consumption in England: importance of average drinker. *BMJ* 1997;314:1164-8.
28. **Saitz R, Lepore MF, Sullivan LM, Amaro H, Samet JH.** Alcohol abuse and dependence in Latinos living in the United States: validation of the CAGE (4M) questions. *Arch Intern Med* 1999;159:718-24.
29. **Moreira LB, Fuchs FD, Moraes RS, Bredemeier M, Cardozo S, Fuchs SC et al.** Alcoholic beverage consumption and associated factors in Porto Alegre, a southern Brazilian city: a population-based survey. *J Stud Alcohol* 1996;57:253-9.
30. **Masur J, Monteiro MG.** Validation of the "CAGE" alcoholism screening test in a Brazilian psychiatric inpatient hospital setting. *Braz J Med Biol Res* 1983;16: 215-8.
31. **Steinweg DL, Worth H.** Alcoholism: the keys to the CAGE. *Am J Med* 1993;94:520-3.
32. **Mayfield D, McLeod G, Hall P.** The CAGE questionnaire: validation of a new alcoholism instrument. *Am J Psychiatry* 1974;131:1121-3.
33. **Gliem JA, Gliem RR.** Calculating, interpreting, and reporting Cronbach's alpha coefficient for likert. type scales. Presented at the Midwest research to practice conference in adult, continuing, and community education. The Ohio State University, Columbus, OH, October 8-10, 2003. Disponible en [www.alumniosu.org/midwest/midwest%20papers/Gliem%20&%20Gliem-Done.pdf](http://www.alumniosu.org/midwest/midwest%20papers/Gliem%20&%20Gliem-Done.pdf). Acceso el 10 de mayo de 2004.
34. **Mora J, Natera G.** Expectativas, consumo de alcohol y problemas asociados en estudiantes universitarios de la ciudad de México. *Salud Pública de Méx* 2001;43: 89-96.
35. **Goldman S, Brown SA, Christiansen BA.** Expectancy theory: thinking about drinking. En: Blane HT, Leonard KE, editors. *Psychological theories of drinking and alcoholism*. New Cork: Guilford Press; 1987. p.181-226.
36. **Brown S, Christiansen BA, Goldman MS.** The alcohol expectancy questionnaire: an instrument for the assessment of adolescent and adult alcohol expectancies. *J Stud Alcohol* 1987;48:483-91.
37. **Sackett DL, Haynes RB.** The architecture of diagnostic research. En: Knottnerus JA, editor. *The evidence base of clinical diagnosis*. London: BMJ Publishing Group; 2002. p.19-38.