

Validación del Cuestionario de Salud General de Goldberg (versión 28 ítems) en consultas de Atención Primaria

Validity of the Goldberg's General Health Questionnaire (28 items version) in Primary Care settings

A. RETOLAZA BALSATEGUI, A. MOSTAJO*, J.R. DE LA RICA**, A. DIAZ DE GARRAMIOLA***, J. PEREZ DE LOZA****, I. ARAMBERRI*****, I. MARQUEZ*****

RESUMEN

Se realiza un estudio para medir la validez del cuestionario GHQ-28 de Goldberg, comparándolo con la entrevista psiquiátrica estructurada PSE-ID, 9.^a edición. Se expone el método utilizado en la selección de 150 pacientes de ambos sexos, todos provinientes de consultas de Atención Primaria, para la entrevista psiquiátrica de validación.

Se analiza el cuestionario en tres órdenes: como test de "screening", como detector de morbilidad psiquiátrica real y como indicador de severidad clínica. Tras comparar los resultados con otros trabajos similares, se concluye con algunas recomendaciones para el uso del GHQ-28.

PALABRAS CLAVE: Validez concurrente. GHQ-28. PSE-ID. Consultas de Atención Primaria. Validez predictiva.

SUMMARY

To measure the validity of Goldberg's GHQ-28 Questionnaire a study is realized, comparing it with the 9th edition PSE-ID structured psychiatric interview. The method used in the selection of 150 patients, both sexes, is exposed, all of them proceeding from primary care settings for the validation psychiatric interview.

The questionnaire is analyzed in three lines: as "screening" test, as real psychiatric morbidity detector and as clinical severity indicator. After comparing the results with another similar works, the task is finished with several recommendations for the use of GHQ-28.

KEY WORDS: Concurrent validity. GHQ-28. PSE-ID. Primary care settings. Predictive value.

INTRODUCCION

El Cuestionario General de Salud (2,3,4) es un instrumento muy extendido para diferenciar, de forma sencilla, probables

Psiquiatra. CSM de Basauri. *Psiquiatra. C. de Salud de Deusto. **Médico. H. Psiq. de Bermeo. ***Psicóloga. CSM «Novia Salcedo» (Bilbao). ****Asistente Social. CSM de Basauri. *****Psicóloga. CSM de Basauri. *****Médico. CSM de Bermeo.

Correspondencia: Ander Retolaza Balsategui. Centro de Salud Mental de Basauri. C/ Nagusia, s/n. 48970 Basauri (Vizcaya).

Fecha de recepción: 17-6-92

Fecha de aceptación: 5-6-93

pacientes psiquiátricos de aquellos considerados como potencialmente normales. Está originalmente diseñado para identificar trastornos mentales no psicóticos en contextos de práctica médica general.

Existen varias versiones de este cuestionario. Las más conocidas son la de 60 ítems (GHQ-60) y la de 28 ítems (GHQ-28), desarrollada ésta a partir de la primera mediante técnicas de análisis factorial. El GHQ-28 consta de cuatro subescalas de siete ítems cada una. La escala A refiere síntomas somáticos; la B, ansiedad e insomnio; la C, disfunción social y la D, depresión severa. Ambas versiones mantienen un buen grado de concordancia, siendo los coeficientes de validez y fiabilidad del GHQ-28 casi tan satisfactorios como los de la versión de 60 ítems (3). Tanto las adaptaciones a nuestro medio (11) como los estudios de validación en nuestro país han dado resultados muy aceptables para ambas versiones en la línea de otros en el extranjero (9,10).

El presente estudio fue desarrollado dentro de una investigación epidemiológica sobre las características de los trastornos psiquiátricos en las consultas de atención primaria correspondientes a nuestro centro. Tiene por objetivo revisar la validez y grado de confianza que merece la versión corta del test (GHQ-28) en un contexto puro de atención primaria (5), en el que apenas ha sido investigado en nuestro país. La más conocida de las validaciones en nuestro entorno fue la realizada por Lobo y cols. entre los pacientes externos de un Hospital Universitario (9).

MATERIAL Y METODO

Los trabajos de campo, tanto para el estudio epidemiológico como para el de validación, se realizaron simultáneamente con pacientes de las consultas de demanda del Centro de Salud de Basauri, durante Febrero y Marzo de 1990 (quince días por mes). Participaron 10 cupos de médicos cuya población asignada es representativa del municipio. Se escogieron 10 días al azar, dentro del periodo a estudio, eligiéndose 5 pacientes por cupo y día también de forma aleatoria. Se incluyeron en el estudio sólo los pacientes que acudían a consulta por motivo propio y con edades comprendidas entre los 18 y los 65 años.

De un total de 500 pacientes participantes en el estudio epidemiológico se obtuvieron 462 con encuestas válidas, que incluían el GHQ-28 y una ficha de datos sociales.

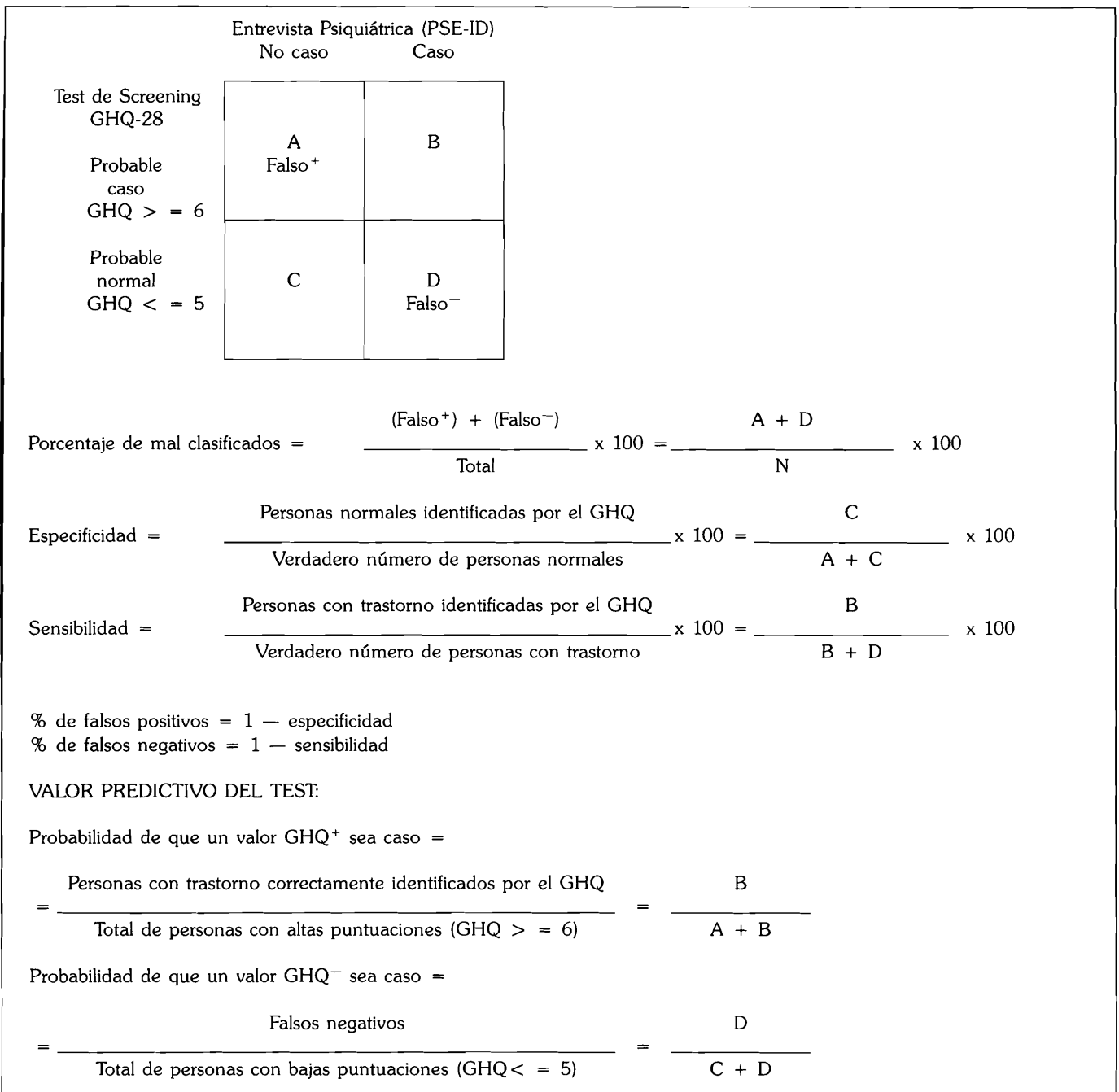


Fig. 1. Descripción de fórmulas utilizadas en el análisis.

Tanto el GHQ-28 como la ficha de variables socioeconómicas fueron cumplimentados en una sala preparada al efecto, donde personal adiestrado resolvía dudas y velaba por la confidencialidad de los resultados. Estos profesionales puntuaban inmediatamente los cuestionarios, seleccionando, según surgían, las personas para la segunda fase. Se escogieron, sucesivamente, los cuatro primeros varones con GHQ⁺ y los cuatro primeros varones con GHQ⁻, lo mismo para las mujeres, cada uno de los 10 días en que se realizó el estudio. Se añadían finalmente 4 ó 5 personas más por día en previsión de rechazos o sustituyendo a los producidos. Se consideró adecuada una muestra final de 150 perso-

nas para entrevista psiquiátrica de validación, esperando una prevalencia en torno al 25% y un 10% de rechazos. Para escoger esta muestra se observaron dos criterios de selección: 1. Obtener un número similar de hombres y mujeres. 2. Obtener un número similar de probables casos y probables no casos, según el GHQ-28, con una línea de corte 5/6 (p. no caso/p. caso) habitualmente utilizada y recomendada.

El sistema PSE-ID

Entre las posibles entrevistas psiquiátricas, se decidió utili-

zar el PSE-ID (Present State Examination, 9.ª edición) (17,18). Sus resultados pueden ser analizados mediante el programa informático CATEGO y el ID (Índice de Definición). Dado que se trata de una entrevista que requiere personal cualificado con entrenamiento en su uso y que el acceso al programa CATEGO está limitado, quizá convenga detenerse en un análisis detallado de este instrumento:

1. Se trata de una entrevista estandarizada que ha demostrado su eficacia en la investigación epidemiológica en psiquiatría (14,15). Originariamente diseñada para el ámbito hospitalario, su criterio de caso resulta más restrictivo que el de otras entrevistas psiquiátricas, por lo que los resultados obtenidos, aunque algo más bajos porcentualmente, son de gran solidez. No obstante presenta algunos inconvenientes como su baja sensibilidad para demencias, psicosis defectuales crónicas, alcoholismo y toxicomanías. Requiere datos actuales referidos al último mes.

2. Consta de 140 ítems, con 54 preguntas obligatorias y un conjunto de preguntas opcionales. Tiene un sistema de marcado directo (puntuaciones posibles: 0, 1, 2, 8, 9), con anulación de preguntas en caso de respuestas negativas. Se estructura en 5 áreas: Neurosis, Cuadros Borderline, Psicosis, Cuadros Psico-orgánicos y Evaluación objetiva. Entre todas ellas incluyen 20 secciones. Se trata, pues, de una entrevista relativamente larga, pero buena para clasificar e investigar. Existen versiones reducidas derivadas de ella.

3. El Índice de Definición (ID) es un marcador de severidad psicopatológica. Se basa en el número, tipo y gravedad de los síntomas que aparecen tras la entrevista PSE. El sistema ID establece 8 grados de definición de síntomas, con 4 niveles bajo umbral o "no caso", que comprenden desde el 1 (no síntomas) hasta el 4 (síntomas específicos) y otros 4 niveles por encima del umbral o "casos", desde el 5 (Borderline) en creciente severidad hasta el 8. A partir del nivel 5 los trastornos están suficientemente bien definidos como para ser sometidos a una tentativa diagnóstica, por el sistema CATEGO, en términos de la Clasificación Internacional de Enfermedades (ICD-9). Con un menor nivel de definición, el sistema elabora también las llamadas "CATEGO Subclases y Clases", que constituyen aproximaciones a síndromes específicos, pero sin diagnóstico definitivo.

Como era de esperar una mayor proporción de mujeres, tanto en la muestra total de partida como entre el grupo GHQ positivo, se mantuvieron todos los criterios de selección para hombres y probables casos los días necesarios hasta igualar al máximo todos los grupos. Las entrevistas PSE se desarrollaron a razón de unos 20 pacientes por día. En ningún caso transcurrieron más de 3 días entre la realización del GHQ-28 y la del PSE.

La entrevista psiquiátrica la realizaron 5 personas (dos psiquiatras, un médico especializado, una psicóloga y una asistente social) ciegas a todo el proceso anterior. Estos entrevistadores fueron previamente entrenados en el PSE por un psiquiatra experimentado en su uso durante 50 horas, en las que se desarrolló un importante grado de confiabilidad en las puntuaciones mediante entrevistas ensayo a pacientes reales.

En el estudio de significación estadística se utilizó la prueba "Chi cuadrado" de Pearson, aceptando un valor máximo

de $\alpha = 0,05$. Para el cálculo de correlaciones se utilizó el coeficiente de Spearman. En el análisis del GHQ-28 se aplicó el método utilizado por Tarnopolsky y cols. (13) descrito en la figura 1.

RESULTADOS

Para alcanzar la cifra de 150 entrevistas PSE fue necesario convocar a 195 personas ya que se produjo un 23% de rechazos. Dado que este grupo se seleccionaba sucesivamente, a medida que se iba administrando el GHQ-28, no hubo problemas en encontrar sustitutos de parecidas características. Entre las 45 personas que rehusaron la entrevista psiquiátrica había 28 mujeres (15 de ellas GHQ ≥ 6) y 17 varones (11 de ellos GHQ ≥ 6). No se encontraron diferencias estadísticamente significativas en cuanto a la distribución por sexo y valores GHQ en este grupo ($p > 0,1$). Los motivos alegados fueron comprobados, siendo en su mayoría enfermedad física y ocupaciones diversas. En 12 ocasiones no se pudo averiguar la causa de rechazo.

La muestra final de 150 personas que pasaron a entrevista psiquiátrica quedó así constituida por 74 hombres (36 GHQ⁺; 38 GHQ⁻) y 76 mujeres (38 GHQ⁺; 38 GHQ⁻). Un total de 32 personas, 21,3% del total, (12 hombres y 20 mujeres) superan el umbral de caso (ID = 5-8) y tienen un diagnóstico ICD-9 asignado. Los diferentes diagnósticos encontrados se detallan en la Tabla I.

Hay 16 personas, 10,7% del total (5 hombres y 11 mujeres), en el nivel de síntomas por debajo del umbral (ID = 4), 9 de ellas presentan estados de ansiedad (CATEGO clase, A⁺), otras 5 síntomas ansioso-depresivos (CATEGO clase, N⁺) y sólo 2 presentan síntomas inespecíficos.

TABLA I

DIAGNOSTICOS SEGUN SEXO Y RESULTADO GHQ-28

Catego clases	ICD-9	Hombres		Mujeres	
		<=5	>=6	<=5	>=6
A ⁺	Neur. Ansiedad	—	2	—	1
A ⁺	Neur. Fóbica	—	1	1	2
N ⁺	Neur. Depresiva	—	6	2	9
M ⁺	PMD, tipo Manía	—	—	0	2
R ⁺	PMD, tipo Depr.	—	2	1	1
O?X	Esquiz. y Psicosis	1	—	—	1
<i>Total</i>		1	11	4	16
			37,5%		62,5%

Análisis de validación del GHQ-28

Las cuatro subescalas del GHQ-28 correlacionan entre sí positivamente en diversos grados y en mayor medida con el conjunto del test. Respecto al PSE encontramos también una correlación positiva aunque no muy alta, tanto para las

subescalas como para el valor total del GHQ. La subescala B (ansiedad e insomnio) es la de mejores resultados frente al PSE, obteniendo ella sola casi tan buena correlación como el total del GHQ. Se advierte un escaso grado de concordancia cuando se intentan comparar las puntuaciones del GHQ con la severidad clínica, medida con el Índice de Definición. Para estos resultados observar las Tablas II, III y IV.

TABLA II

INTERCORRELACIONES ENTRE SUBESCALAS Y PUNTUACION TOTAL GHQ-28

	Escala B	Escala C	Escala D	Total GHQ-28
Escala A	0,53	0,39	0,32	0,78
Escala B	—	0,45	0,51	0,86
Escala C	—	—	0,52	0,70
Escala D	—	—	—	0,64

TABLA III

CORRELACION ENTRE SUBESCALAS Y GHQ-28 TOTAL CON PSE

	Escala A	Escala B	Escala C	Escala D	GHQ-28
PSE total	0,35	0,50	0,36	0,44	0,53

TABLA IV

CORRELACION ENTRE SUBESCALAS. GHQ-28 E INDICE DE DEFINICION (ID) POR NIVELES

Índice de Definición			Esc. A	Esc. B	Esc. C	Esc. D	Total GHQ-28
PSE							
No síntomas	1						
	2		0,18	0,24	—	—	0,27
Síntomas inespecíficos	3						
	4		—	—	—	—	—
Síntomas específicos Borderline	5		—	—	—	—	—
	6		—	—	—	—0,34	—0,35
Casos	7		—	—	—	—	—
	8		—	—	—	—	—

Nota: Sólo se incluyen aquellos con significación estadística ($p < 0,05$).

Dado que el PSE-ID suele producir un más alto nivel umbral para la identificación de casos que el CIS (Clinical Interview Schedule), usado en la validación española original del GHQ-28 (8,9,14), pareció oportuno observar también el comportamiento del test utilizando otras líneas de corte. La Tabla V muestra, en números absolutos, estos resultados en función del sexo y de los normales y casos hallados en la entrevista psiquiátrica. Los porcentajes derivados se pueden ver en la Tabla VI.

Observamos, para la línea de corte 5/6, una sensibilidad igual o superior al 80% en ambos sexos, mientras que la especificidad cae a valores en torno al 60%. En consonancia con ésto el valor de predicción negativa alcanza altos valores, mientras que el de predicción positiva es más pobre, sobre todo para los hombres. Si elevamos la línea de corte se produce una progresiva reducción del porcentaje de mal clasificados, aumentando algo el valor predictivo positivo y manteniéndose el negativo en cifras muy aceptables. Esta relativa mejora de la eficiencia del test se produce, a costa de una pérdida de sensibilidad, sin elevarse demasiado el porcentaje de falsos negativos. La línea de corte 7/8 resulta en conjunto la más equilibrada, según nuestros resultados, con el inconveniente de una importante diferencia en la sensibilidad del GHQ-28 para ambos sexos.

Puesto que el criterio de caso en el PSE parece tener un umbral más alto que el usado en el CIS, puede esperarse que aquél clasifique como "normales" a un cierto número de personas cuyos síntomas hubieran bastado para considerarlos "caso" en los trabajos en los que el CIS fue la entrevista

TABLA V

LINEAS DE CORTE GHQ-28 EN REFERENCIA A ENTREVISTA PSIQUIATRICA PSE-ID

GHQ-28		Hombres (n = 74) E. Psiquiátrica (PSE-ID)		Mujeres (n = 76) E. Psiquiátrica (PSE-ID)		Muestra total (n = 150) E. Psiquiátrica (PSE-ID)	
		Normal (ID 1-4) n = 62	Caso (ID 5-8) n = 12	Normal (ID 1-4) n = 56	Caso (ID 5-8) n = 20	Normal (ID 1-4) n = 118	Caso (ID 5-8) n = 32
		Línea de corte 4/5	Valores+ n = 84	30	11	25	18
	Valores- n = 66	32	1	31	2	63	3
Línea de corte 5/6	Valores+ n = 74	25	11	22	16	47	27
	Valores- n = 76	37	1	34	4	71	5
Línea de corte 6/7	Valores+ n = 63	20	11	18	14	38	25
	Valores- n = 87	42	1	38	6	80	7
Línea de corte 7/8	Valores+ n = 53	14	10	16	13	30	23
	Valores- n = 97	48	2	40	7	88	9
Línea de corte 8/9	Valores+ n = 46	14	8	12	12	26	20
	Valores- n = 104	48	4	44	8	92	12

TABLA VI

VALORES ESTADISTICOS DEL GHQ-28 PARA DIFERENTES PUNTOS DE CORTE

Líneas de corte GHQ	4/5			5/6			6/7			7/8			8/9		
	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T
Sensibilidad	91,7	90,0	90,6	91,7	80,0	84,4	91,7	70,0	78,1	83,3	65,0	71,9	66,7	60,0	62,5
Especificidad	51,7	55,4	53,4	59,7	60,7	60,2	67,7	67,9	67,8	77,4	71,4	74,6	77,4	78,6	78,0
V. Pred. Positivo	26,9	41,9	34,5	30,6	42,1	36,5	35,5	43,8	39,7	41,7	44,8	43,4	36,4	50,0	43,5
V. Pred. Negativo	96,9	93,9	95,5	97,4	89,5	93,4	97,7	86,4	92,0	96,0	85,1	90,7	92,3	84,6	88,5
Mal clasificados	41,9	35,5	38,7	35,1	34,2	34,7	28,4	31,6	30,0	21,6	30,3	26,0	24,3	26,3	25,3

TABLA VII

COEFICIENTES DE VALIDEZ DEL GHQ-28 FRENTE A "UMBRAL BAJO" PSE-ID ("NORMAL": ID 1-3; "CASO": ID 4-8)

GHQ línea de corte 5/6	Entrevista psiquiátrica PSE-ID					
	Hombres n = 74		Mujeres n = 76		Muestra Total n = 150	
	"Normal" (ID 1-3)	"Caso" (ID 4-8)	"Normal" (ID 1-3)	"Caso" (ID 4-8)	"Normal" (ID 1-3)	"Caso" (ID 4-8)
Valores ⁺ (GHQ = 6-28)	21	15	14	24	35	39
Valores ⁻ (GHQ = 0-5)	36	2	31	7	67	9
Sensibilidad	88,2		77,4		81,3	
Especificidad	63,2		68,9		65,7	
V. Pred. Positivo	41,7		63,2		52,7	
V. Pred. Negativo	94,7		81,6		88,2	
Mal clasificados	31,1		27,6		29,3	

ta psiquiátrica de referencia (14). En tal caso, si bajamos el umbral de "normalidad" de PSE-ID hasta el nivel 3/4 (Normal: ID 1-3; Caso; ID 4-8) deberíamos obtener un mejor rendimiento general del GHQ-28. La Tabla VII muestra que, en estas condiciones, la especificidad y el valor de predicción positiva mejora, mientras la sensibilidad y el valor de predicción negativa se mantienen aceptables. Con ligeras diferencias cuantitativas, estas variaciones se producen en la misma dirección para ambos sexos.

Casos mal clasificados por el GHQ-28

En conjunto la proporción de casos mal clasificados es superior al 25% en todas las líneas de corte y mayor cuanto más bajas sean éstas. Por sexo, la línea 5/6 resulta la más equilibrada en este aspecto, siendo mayor la proporción de mujeres mal clasificadas respecto a la de hombres para las líneas de corte más altas.

Considerando el corte 5/6 hay 47 personas (25 hombres y 22 mujeres) clasificadas como "falso positivo" en nuestro estudio. Se producen 5 "falsos negativos" (1 hombre y 4 mujeres).

A fin de establecer si existe algún factor asociado que contribuya a la aparición de "falsos positivos" resulta útil comparar este grupo con el de "verdaderos negativos". Hemos encontrado mayor número de personas con quejas de molestias físicas o referidos como pacientes orgánicos por su médico de familia entre los "falsos positivos", pero no ha sido posible encontrar ninguna diferencia significativa, respecto a las variables socio-económicas estudiadas, entre ambos grupos.

El grupo de "falsos negativos" está formado por un toxímano con patología orgánica severa, una paciente esquizofrénica de larga evolución, compensada y con escasa producción sintomática, que había acudido a su médico por enfermedad banal, una enferma depresiva grave afecta de ferropenia y dos mujeres con trastorno ansioso-depresivo crónico, una de ellas con posible hipertiroidismo.

Si estudiamos la probabilidad de que una puntuación GHQ

negativa sea caso, esto es, la probabilidad de dar "falso negativo", observamos que ésta es mayor, para ambos sexos, cuanto más alta sea la línea de corte. Pero la tendencia es notoriamente más acusada entre las mujeres en todas las situaciones observadas, resultando siempre mayor la proporción de falsos negativos entre ellas.

DISCUSION

Cabría pensar que la relativamente alta proporción de personas que rechazaron la entrevista psiquiátrica introdujera un sesgo que desvirtuara nuestros resultados. Sobre todo si los casos perdidos afectaran más a un sexo que a otro, o a la distribución valores del GHQ-28 por encima y debajo del punto de corte de referencia. El estudio de las características de este grupo nos permite excluir, en grado razonable, esta posibilidad ya que la distribución de valores GHQ-28 por sexo no es excesivamente diferente de la del grupo que pasó a realizar el PSE y, en todo caso, no resulta estadísticamente significativa. Los motivos alegados de rechazo que se pudieron constatar fueron la mayoría y apuntan en direcciones con escaso riesgo de sesgar la muestra a estudio.

El GHQ-28 como test de "screening"

La sensibilidad, o proporción de casos reales correctamente identificados, es considerado el valor estadístico de mayor interés en el "screening". Mide lo que se denomina la "Utilidad del test". Complementariamente el "Costo del test" viene determinado por la proporción de falsos negativos (1-sensibilidad). La especificidad y la proporción de casos mal clasificados constituyen también elementos a ser tenidos en cuenta en este apartado (13).

La sensibilidad y la especificidad dependen de la proporción GHQ⁺/GHQ⁻ en la muestra a estudio. Algunos autores consideran asimismo importante tener en cuenta la proporción hombres/mujeres, al haberse reportado algunas diferencias de interés, en cuanto a sensibilidad y/o especifi-

cidad según sexo (13). Por todo ello, las muestras igualadas al 50%, según resultados en el GHQ y sexo, resultan muy adecuadas en los estudios de validación, confiriendo gran valor a los resultados.

Podemos confirmar un buen nivel de sensibilidad del GHQ-28 utilizando la línea de corte 5/6 recomendada por Lobo y cols. en la validación española (9). Existe, no obstante, una diferencia de más de 10 puntos porcentuales en la sensibilidad considerada por sexo que debe ser anotada como importante y que se confirma en todas las líneas de corte superiores. A igual proporción $\text{GHQ}^+/\text{GHQ}^-$ el test recoge con mayor cuidado los casos reales presentes entre la población masculina. En nuestra muestra el sexo parece afectar a la respuesta al test, aspecto ya anotado en otros trabajos que nos preceden (1,13).

Aunque la especificidad obtenida resulta baja no consideramos ésto un serio inconveniente a la hora de juzgar la capacidad de "screening" del test. Primero, porque lo importante en la función de "screening" es no perder casos reales y éstos fueron muy pocos; segundo, porque los valores de especificidad que hemos encontrado para las diversas líneas de corte están en consonancia con los obtenidos en otros estudios en que se ha utilizado el PSE-ID como entrevista psiquiátrica de referencia (14). De todas formas, la línea de corte 4/5, recomendada por Goldberg (3,4), resulta excesivamente baja a tenor de nuestros resultados, que son mejores si los referimos a la línea 7/8.

Comparando con otros trabajos (9,13), la proporción de casos mal clasificados parece excesivamente alta en nuestro estudio. Pensamos que en ello pueden influir las características del PSE-ID, cuyo criterio de caso psiquiátrico produce este efecto general en relación con el CIS (14). Esta apreciación se confirma cuando vemos que el grupo se compone en su gran mayoría de falsos positivos. No hemos podido establecer una presencia significativa de personas de edad avanzada, viudos o desempleados entre éstos, a diferencia de otras investigaciones (14,15). Aunque no hemos estudiado factores como el estrés transitorio o acontecimientos adversos de la vida, la presencia de enfermedad somática, cargas sociales y problemas domésticos resultaba muy habitual entre la población estudiada. Se sabe que estas circunstancias suelen incrementar el número de falsos positivos (9,14).

Respecto a los falsos negativos nuestros resultados confirman hallazgos anteriores (7,9). Suelen darse entre personas con trastornos psiquiátricos crónicos, habituados en cierto sentido a sus síntomas y con una actitud defensiva ante ellos. El alcoholismo y demás drogadicciones se caracterizan también por esta tendencia defensiva y negadora de síntomas, que pueden, sin embargo, ser establecidos tras una entrevista experta.

El GHQ-28 como indicador de morbilidad

Dada la facilidad de administración del test y las dificultades inherentes a la realización de entrevistas psiquiátricas estandarizadas, se ha venido introduciendo al GHQ-28, en algunos estudios epidemiológicos, como referente único de

posible morbilidad psiquiátrica (12). Resulta entonces de interés calibrar la capacidad del test en este orden, comparándolo con otros instrumentos de detección, particularmente las mencionadas entrevistas psiquiátricas.

El estadístico más relevante en este apartado es el valor de predicción, o probabilidad de una puntuación dada en el GHQ-28 de identificar un caso psiquiátrico real. El valor predictivo positivo mide esta probabilidad cuando la puntuación del GHQ-28 está por encima de línea de corte. La probabilidad de dar falsos negativos es el equivalente cuando la puntuación del GHQ-28 está bajo la línea de corte. Esta segunda medida es diferente del valor predictivo negativo, en cuanto que lo que determina es la probabilidad de ser caso psiquiátrico que tienen las puntuaciones por debajo de la línea de corte. Resulta, pues, de mayor utilidad en este apartado.

Los valores de predicción no dependen de la proporción relativa $\text{GHQ}^+/\text{GHQ}^-$ en la muestra y son constantes para una determinada línea de corte. En cambio varían con éstas y, también, en función de la prevalencia real encontrada, esto es, en función de los casos confirmados por la entrevista psiquiátrica. Si la prevalencia es muy alta o muy baja los resultados del test son peores. Se considera que éste tiene un rendimiento óptimo en los tramos intermedios de prevalencia real (3,6).

Nuestros resultados indican que los valores de predicción positiva mejores para ambos sexos se sitúan por encima de las líneas de corte habitualmente utilizadas. En todo caso no parecen lo suficientemente satisfactorios como para confiar en el uso del test como indicador de morbilidad real, al ser inferiores siempre a valores del 50%. Hemos de considerar, una vez más, el efecto "umbral alto" del PSE para una más ajustada interpretación y comparación de estos datos.

Merece una cuidadosa valoración el hecho de que la capacidad de predicción del test resulte menor para los hombres. Esto se produce porque la entrevista psiquiátrica ha establecido casi el doble de casos reales femeninos que masculinos, a pesar de que el número de los seleccionados por el GHQ como positivos era casi igual en ambos sexos. Sin embargo existe una reconocida tendencia de las mujeres a referir más viva y abiertamente sus síntomas que podría suponer una cierta minusvaloración en la entrevista psiquiátrica de casos reales masculinos, dada su menor expresividad, en contra de sus altas puntuaciones en el test (1,13).

La prevalencia total encontrada, según la entrevista psiquiátrica (21,3%), está dentro de un rango intermedio y no parece haber contribuido a disminuir la eficiencia global del GHQ; sin embargo las diferencias, ya señaladas, de prevalencia por sexo pueden explicar el diverso comportamiento del test para hombres y mujeres, puesto que la proporción de casos reales masculinos sí resultó baja, respecto al total de la muestra.

El GHQ-28 como indicador de severidad

Así como las bajas puntuaciones del GHQ predicen con bastante fiabilidad los no casos, resulta razonable esperar que cuanto más altas sean las puntuaciones mayor será la seve-

ridad sintomática real encontrada. La existencia de cuatro subescalas diferenciadas debe añadir, además, alguna información relevante (16).

Desde una perspectiva general el GHQ se muestra como un pobre indicador de severidad de síntomas y carece de precisión diagnóstica (13). Nuestros resultados muestran que las puntuaciones medias y altas del GHQ, por encima de la línea de corte, no se ajustan adecuadamente a la severidad clínica encontrada tras la entrevista psiquiátrica. No parece aconsejable pretender obtener información sobre la gravedad de los pacientes con el GHQ-28.

El hecho de que la subescala B concuerde mejor que las otras tanto con el valor total del GHQ como con los resultados de la entrevista psiquiátrica ya ha sido mencionado por el propio Goldberg (4). La ansiedad y el insomnio se comportan como fenómenos nucleares, comunes a la mayoría de los trastornos psíquicos.

CONCLUSIONES

Se puede confirmar lo ya destacado por otros trabajos que nos preceden, concluyendo que el GHQ-28 resulta un buen test de "screening", pero sería adecuado ajustar la línea de corte a las características de la muestra estudiada a fin de ob-

tener mejores resultados (6,13). Es menos fiable como indicador de morbilidad, pudiendo utilizarse en este sentido siempre que podamos recalibrar su sensibilidad y especificidad también en función de cada muestra concreta. No resulta, sin embargo, un buen predictor de gravedad clínica.

Conviene, en consecuencia, realizar entrevistas psiquiátricas estandarizadas con parte de cada muestra a fin de llevar a cabo los ajustes convenientes. El valor del GHQ-28 reside en que ayuda a reducir sensiblemente el número de éstas, aspecto importante ante muestras muy grandes.

Parece de interés profundizar en el conocimiento de las diversas subescalas, a fin de evaluar si realmente son útiles para suministrar una información más matizada en cada una de las áreas que pretenden acotar.

AGRADECIMIENTOS

Este trabajo debe mucho al asesoramiento del Profesor J.L. Vázquez Barquero. Gracias también a los médicos de familia del Centro de Salud de Basauri, a Javier Ballesteros del Dpto. de Psiquiatría de la UPV y a Luis C. Cubas de la Universidad de Deusto. El Dpto. de Sanidad del Gobierno Vasco becó el proyecto.

BIBLIOGRAFIA

1. Briscoe ME, Vázquez-Barquero JL, Williams P, Díez Manrique JF, Peña C. Sex Differences in the Differentiation of Psychiatric Symptomatology. *British Journal of Psychiatry* 1989; 154: 364-367.
2. Goldberg D. The detection of Psychiatric Illness by Questionnaire. Maudsley Monograph. Oxford University Press. London 1972; 21.
3. Goldberg D. (1978) Manual of the General Health Questionnaire. NFER Publishing Company. Windsor 1978.
4. Goldberg D, Hillier VF. (1979) A scaled version of the General Health Questionnaire. *Psychological Medicine* 1979; 9: 139-145.
5. Goldberg D, Huxley P. Mental Illness in the Community. The Pathway to Psychiatric Care. Tavistock Publications. London & New York 1980.
6. Hobbs P, Ballinger CB, Smith AHW. Factor Analysis and Validation of the General Health Questionnaire in Women: A General Practice Survey. *British Journal of Psychiatry* 1983; 142: 257-264.
7. Lobo A. "Screening" de trastornos psíquicos en la práctica médica. Secretariado de Publicaciones. Universidad de Zaragoza 1987.
8. Lobo A, Gómez Burgada F, Pérez Echeverría MJ, Miranda M, Clare AW. Estudio de la entrevista psiquiátrica estructurada CIS en pacientes españoles. *Comunicación Psiquiátrica* 1984; 11: 43-60.
9. Lobo A, Pérez Echeverría MJ, Artal J. (1986) Validity of the scaled version of the General Health Questionnaire (GHQ-28) in a Spanish population. *Psychological Medicine* 1986; 16: 135-140.
10. Muñoz PE, Vázquez JL, Pastrana E, Rodríguez F, Oneca C. (1978). Study of the Validity of Goldberg's 60 item GHQ in its Spanish version. *Social Psychiatry* 1978; 13: 99-104.
11. Muñoz PE, Vázquez JL, Rodríguez Insausti F, Pastrana E, Varo J. Adaptación española del General Health Questionnaire (GHQ) de DP Goldberg. (Un método de identificación de casos psiquiátricos en la comunidad). *Archivos de Neurobiología* 1979; 42, 2: 139-158.
12. Padierna Acero JA, Gastián Sáenz F, Díaz López P, Etxbeste Antón A, Martínez López E, Nieto Pérez A, Oronoz Badiola P, Sopeña Robres V. (1988). La morbilidad psiquiátrica en Atención Primaria: Detección y derivación por el médico de familia. *Revista AEN* 1988; Vol VIII, 24: 21-29.
13. Tarnopolsky A, Hand DJ, McLean EK, Roberts H, Wiggins RD. Validity and Uses of a Screening Questionnaire (GHQ) in the Community. *British Journal of Psychiatry* 1979; 134: 508-15.
14. Vázquez-Barquero JL, Díez-Manrique JF, Peña C, Quintanal RG, Labrador López M. Two Stage Design in a Community Survey. *British Journal of Psychiatry* 1986; 149: 88-97.
15. Vázquez Barquero JL, Díez-Manrique JF, Peña C, Aldama J, Samaniego Rodríguez C, Menéndez Arango J, Mirapeix C. A community mental health survey in Cantabria: a general description of morbidity. *Psychological Medicine* 1987; 17: 227-241.
16. Vázquez-Barquero JL, Williams P, Díez-Manrique JF, Lequerica J, Arenal A. The factor structure of the GHQ-60 in a community sample. *Psychological Medicine* 1988; 18: 211-218.
17. Wing JK, Cooper JE, Sartorius N. Measurement and Classification of Psychiatric Symptoms. Cambridge University Press. Cambridge 1974.
18. Wing JK, Sturt E. The PSE-ID-CATEGO System: A supplementary Manual. Institute of Psychiatry: London 1978.