

---

## FUNDAMENTOS TEORICOS Y METODOLOGICOS DE LA CLASIFICACION EN PSICOPATOLOGIA

---

CARMELO VÁZQUEZ

Uno de los problemas básicos de la actividad científica en Psicopatología es el de la clasificación. En torno a él se articulan algunos de los retos y controversias más importantes de dicha disciplina. El problema de la clasificación en Psicopatología ha suscitado polémicas apasionantes, pues bajo su rúbrica subyacen problemas tan esenciales como la distinción entre normalidad y anormalidad o la covariación de síntomas.

Por otro lado, la discusión de las bases epistemológicas de la clasificación es un debate necesario, y apenas frecuentado, en Psicopatología. Muchas de las descalificaciones tradicionales contra las taxonomías se han efectuado en base a los sistemas nosológicos existentes en ese momento y desde ahí se ha generalizado dicha crítica hasta incluso cuestionar la propia legitimidad de la *actividad*

clasificatoria. Así, por ejemplo, las adecuadas críticas sobre la ineficacia de las clasificaciones para la planificación terapéutica (Matarazzo, 1983) no se han basado en razones teóricas, sino meramente empíricas. Por tanto, ha de distinguirse claramente entre lo que son las «bases teóricas de la clasificación» y los sistemas clasificatorios que se han utilizado *de facto*. Si no se hace esto, se corre el riesgo, como suele ser habitual, de plantear críticas obsoletas y estereotipadas respecto a la clasificación.

### EL PROBLEMA DE LA CLASIFICACION

#### Bases teóricas de la clasificación

En el proceso constitutivo de *cualquier* disciplina científica, la actividad

clasificatoria es esencial, puesto que el objetivo de ésta es descubrir algún orden en los elementos objeto de su estudio. En otras palabras, «... la búsqueda de una taxonomía ideal descansa en la asunción de un orden intrínseco en el universo» (Millon, 1987, p. 14).

Es necesario, sin embargo, efectuar una distinción semántica previa que nos permita distinguir la multiplicidad de significados de este término. De modo más concreto, en la investigación sobre la clasificación se distinguen tres áreas diferentes (Simpson, 1961; Mezzich, 1984a; Millon, 1987):

1. *Taxonomía*. Por este término, procedente de las ciencias biológicas, se entiende el estudio teórico de los problemas clasificatorios de un objeto de estudio concreto, sin mención a clasificaciones concretas. La taxonomía puede entenderse, pues, como la teoría de la clasificación.

2. *Clasificación*. Básicamente consiste en la formación de grupos o sujetos en categorías homogéneas. El conjunto organizado de grupos forma lo que se conoce como *nomenclatura*, *sistema clasificatorio*, o, en el caso de las enfermedades, *nosología*.

3. *Identificación*. Es el proceso de asignar sujetos —mediante una serie de reglas o principios— a algunos de los grupos que conforman la nomenclatura. El proceso de identificación es lo que suele denominar *diagnóstico*, aunque también, a veces, se le denomina *clasificación*. Así pues, por «clasificación» entendemos tanto el *proceso* como el *producto* de configuración de grupos homogéneos.

Aunque en muchas ocasiones se concibe la clasificación psicopatológica como algo artificialmente impuesto (por

ejemplo, Szasz, 1961; Kanfer y Saslow, 1965), creemos que, en realidad, la actividad clasificatoria es una expresión natural del funcionamiento cognitivo humano. En efecto, las personas ha de valerse, para funcionar adecuadamente en un medio complejo, de ciertos principios de ordenamiento del flujo incesante de información que continuamente reciben. Parece probado que los seres humanos, a modo de «científicos ingenuos», emplean cotidiana y constantemente principios clasificatorios y diagnosticadores (Vázquez, 1985, 1986).

Esta perspectiva «naturalista» nos parece sumamente sugerente como punto de partida de la actividad diagnóstica (véase también Spitzer y Wilson, 1982). Obviamente, un posible paralelismo entre la función cognitiva de los seres humanos y el funcionamiento interno de las ciencias, no puede justificar por sí solo la bondad de los sistemas clasificatorios. Sin embargo, creemos que se puede apelar legítimamente a esta correspondencia *funcional* para, cuando menos, resaltar el sustrato lógico de la clasificación.

¿Cuántas personas y quiénes padecen un determinado trastorno mental? ¿Cómo se deben asignar los recursos económicos en el estudio de estas alteraciones? ¿Cómo se establecen las prioridades en la política sanitaria de un país? ¿Cuál es la probable evolución de un cuadro psicopatológico? ¿Quién se beneficia de las terapias?... Las clasificaciones sirven como el auxiliar básico para intentar dar respuesta a estas importantes preguntas. Así, por ejemplo, la investigación epidemiológica depende manifiestamente de la existencia de criterios diagnósticos.

Por otro lado, las clasificaciones en Psicopatología obedecen, de modo simultáneo, a fines diversos y, a menudo, irreconciliables (científicos, estadísticos,

organizativos, etc.). En definitiva, además del obvio objetivo de crear una herramienta que recoja el estado del conocimiento sobre la conducta anormal y que sirva como catalizador de hipótesis de trabajo, las clasificaciones cubren también objetivos estadísticos e incluso, más prosaico aún, objetivos administrativos. Esta diversidad de condicionantes hace bastante improbable que se pueda llegar alguna vez a un sistema clasificatorio único y multipropósito (Meehl, 1987; Eysenck et al., 1983).

En cualquier caso, parece improbable cualquier tipo de actividad científica sin apelar a un sistema clasificatorio. Incluso desde aquellas posiciones desde las que se ha hecho gala de una ausencia de clasificaciones (por ejemplo, el conductismo), inevitablemente se han empleado criterios clasificatorios *implícitos* sobre la conducta anormal (Vázquez, 1985). Esta posición de rechazo frontal no resuelve el problema, pues, como insistiremos a lo largo de este capítulo, una de las fuentes de discrepancias y desacuerdos en el diagnóstico puede proceder de la ausencia de especificación de criterios diagnósticos. En efecto, estas clasificaciones implícitas no poseen el rigor y verificabilidad de las clasificaciones formales, entre otras razones porque existen múltiples métodos de clasificar los fenómenos.

La discusión anterior no debe conducir a la conclusión de que dado que existe una función «categorizadora» cognitiva casi inevitable en los seres humanos, debemos aceptar inevitablemente las clasificaciones disponibles en la actualidad. De hecho, es absolutamente legítimo prescindir, por razones de utilidad o eficacia, de un determinado esquema clasificatorio (por ejemplo, las clasificaciones psiquiátricas al uso) como *guías* de la actividad clínica. Lo único que pretende-

mos es mostrar la viabilidad lógica de las clasificaciones.

En resumidas cuentas, las clasificaciones no son un mero epifenómeno de la actividad científica o un artefacto vergonzante de la misma. Por el contrario, la clasificación tiene un papel epistemológico relevante, pues, en realidad, no supone sino un «conjunto de proposiciones explicativas... que otorgan significado a nuestras experiencias clínicas» (Millon, 1987). Es decir, la clasificación supone la *explicitación* de una serie de ejes y criterios que permiten, en último término, identificar individuos homogéneos respecto a tales criterios.

## Clasificación y realidad

Un punto de partida crucial para comprender la naturaleza del diagnóstico, lo constituye el reconocer que las etiquetas diagnósticas en Psicología no son «realidades» en sí mismas. La «esquizofrenia» o la «depresión» no son algo palpable, tangible, sino una construcción teórica que nos permite comunicarnos a los clínicos. Aunque en las ciencias físicas las clasificaciones suelen tener referentes bastante precisos y directamente observables, incluso en la física subatómica, muchas partículas son también, en realidad «constructos» o conceptos teóricos que permiten explicar ciertos fenómenos experimentales sutiles (Sánchez del Río, 1988). Debe quedar claro que dada la complejidad y versatilidad de la conducta humana, las clasificaciones de ésta tienen un carácter necesariamente *hipotético*. La conducta humana está sujeta a multitud de determinantes que, en muchos casos, apenas son predecibles (Bandura, 1982), lo que hace especialmente difícil su estudio.

En definitiva, desde una perspectiva

epistemológica, las categorías diagnósticas han de entenderse como una serie de *entidades de naturaleza hipotética* que han de revisarse sistemáticamente en base a la investigación empírica que *paralelamente* se desarrolla para su verificación (por ejemplo, Zwick, 1983).

Por último, no conviene olvidar que si bien las categorías psiquiátricas no son entidades reales en sentido estricto, pueden tener consecuencias estigmatizadas nefastas sobre el individuo que las recibe (Szasz, 1961; Rosenhan, 1973; Scheff, 1981). Esta observación, situada en un nivel de análisis más sociológico que epistemológico, supone un elemento crítico de los sistemas diagnósticos que no se debe menospreciar. De hecho, las clasificaciones más recientes, como los RDC (véase el siguiente capítulo), han tenido este argumento en cuenta, de tal modo que su objetivo ha sido confeccionar criterios diagnósticos *conservadores* con el fin de reducir al máximo el número de personas erróneamente diagnosticadas.

## FUNCIONES DE LA CLASIFICACION

¿Para qué sirven las clasificaciones? Un análisis detallado (véase, por ejemplo, Mezzich, 1984 y Blashfield y Draguns, 1976a), permite identificar al menos las siguientes funciones específicas:

1. *Organización de la información clínica.* Una etiqueta clasificatoria debería permitir manejar la información que se dispone del paciente de un modo conciso. La clasificación de un individuo facilita de modo *económico* esta función. Desde este punto de vista las etiquetas de cualquier sistema clasificatorio son elementos que, a modo de *esquemas*, guían la observación y filtraje de los estí-

mulos entrantes y, por ende, sirven para interpretar la realidad (Millon, 1987). La actividad clasificatoria, por tanto, operaría a modo de un esquema cognitivo que permite un procesamiento más eficaz de la información que se recoge del sujeto.

Es evidente que la contrapartida de este proceso es una pérdida importante de información idiosincrática que podría ser relevante para caracterizar al individuo clasificado. Muchos críticos de las nosologías psicopatológicas han centrado sus ataques en este carácter reductor de las *etiquetas*. Asimismo, la etiqueta puede acabar funcionando como un estereotipo que llegue a distorsionar la percepción y manejo de la información sobre el paciente (Rosenhan, 1973).

Una crítica conceptual interesante al diagnóstico tradicional, es que en éste se acentúan básicamente características *negativas* del individuo (Kanfer y Saslow, 1965; Albee, 1970; Korchin, 1976). Esta observación es importante no sólo porque contribuye a los efectos negativos de la etiqueta diagnóstica, sino porque puede *sesgar* la visión del clínico. Es posible que para formular una predicción sobre el curso de una alteración o incluso para efectuar investigaciones etiológicas, sea tan importante analizar los elementos «patológicos» del individuo (recogidos en la etiqueta) como aquellas características adaptativas que posee o conserva. Sin embargo, este último enfoque está casi totalmente ignorado por los clínicos. Las nosologías psiquiátricas se centran casi exclusivamente en resaltar lo patológico sin analizar los elementos positivos del repertorio psicológico o conductual del sujeto. Este sesgo puede entenderse como un resultado de lo que es un proceso cognitivo normal: lo inusual, lo anómalo, es lo que resulta más llamativo en la percepción de cualquier objeto o fenómeno.

La función economizadora del diagnóstico queda bien expresada en la definición que sugieren Robins y Helzer (1986, p. 430): «Los diagnósticos son, como mucho, descripciones imperfectas de la realidad. No expresan plenamente la peculiaridad del individuo y utilizan de modo arbitrario puntos de corte que excluyen a algunas personas que requieren tratamiento e incluso a otras que no lo requieren. Pero los diagnósticos de un sistema bien estructurado proporcionan un medio útil de comunicar sucintamente una buena cantidad de información, aunque de ningún modo toda, sobre la historia del paciente y su inmediato futuro.»

2. *Facilitar la comunicación entre profesionales.* Quizás sea ésta la función que justifica mejor la existencia de los esquemas clasificatorios. Uno de los objetivos de cualquier clasificación es el de proporcionar un marco semántico compartible. Esto es especialmente urgente en Psicopatología y Psiquiatría, pues la coexistencia de diferentes teorías, nombres de variables y procesos, y métodos de medida, hace deseable la existencia de puntos comunes de referencia. El uso de términos comunes es, por sí mismo, un punto de partida crucial para cualquier ciencia.

3. *Predicir el curso clínico y la respuesta al tratamiento.* Uno de los objetivos prácticos más importante de las categorías diagnósticas sería el de proporcionar una estimación del pronóstico del sujeto. No hay que olvidar que el sistema clasificatorio de Kraepelin se fundaba básicamente en el pronóstico del cuadro; esta distinción era lo que servía como base teórica de muchas de sus distinciones diagnósticas.

4. *Seleccionar el tratamiento adecuado.* En una clasificación ideal, habitual-

mente se considera que el diagnóstico puede ser una vía importante en la determinación del tratamiento. Incluso en el estado actual de conocimientos en Psicopatología, la variable «diagnóstica» parece más importante que otras variables (sexo, nivel socioeconómico, raza, etc.) para tomar una determinación terapéutica global (Guze, 1977).

En el tratamiento de las enfermedades orgánicas, la adscripción a una categoría suele ser suficiente para identificar adecuadamente el problema y proponer un *tratamiento*. Es este modelo el que normalmente se asume en la práctica psiquiátrica. Sin embargo, las estrategias de intervención psicológica han de valerse de información adicional, pues se trata de cambiar segmentos más o menos amplios de conductas y para esto necesariamente se requiere formulaciones más complejas y detalladas del problema.

5. *Clarificar la etiología de la alteración.* La selección de grupos diferentes de sujetos facilita la posibilidad de averiguar vías y nexos comunes en el desarrollo de la alteración.

6. *Facilitar el desarrollo teórico.* La clasificación ayuda a orientar la investigación. Por ejemplo, la distinción de dos grupos de esquizofrenia —paranoide y no paranoide— ha demostrado ser de particular relevancia en el descubrimiento de procesos cognitivos diferenciales en ambos subgrupos de sujetos (Mogaro, 1980). Sin embargo, las clasificaciones también pueden limitar la construcción teórica. Como veremos en el siguiente capítulo, a veces se ha señalado que atenerse a las nosologías actuales en realidad puede oscurecer el análisis de los *procesos* fisiológicos y psicológicos implicados en los trastornos mentales.

## EVALUACION DE LOS SISTEMAS CLASIFICATORIOS: FIABILIDAD Y VALIDEZ

Las clasificaciones tienen siempre carácter tentativo y provisional y, en consecuencia, han de someterse, como cualquier otra hipótesis, a un proceso empírico de verificación. Los sistemas nosológicos psiquiátricos apenas han resistido un análisis de este tipo, aunque se han mantenido vigentes por inercia, mimetismo con otras disciplinas, o intereses ajenos a la ciencia.

Tradicionalmente, los sistemas clasificatorios han sido evaluados atendiendo a dos criterios fundamentales: *fiabilidad* y *validez*. En la investigación diagnóstica actual, también se consideran otros dos importantes requisitos que subrayan la eficiencia con se efectúa el diagnóstico: la *sensibilidad* y la *especificidad*. Estos cuatro elementos, en su conjunto, conforman una matriz crucial para contrastar la utilidad de una clasificación dada.

### FIABILIDAD

*Fiabilidad interjueces.* Por fiabilidad se entiende la precisión con que un instrumento dado evalúa lo que está midiendo. Si bien la fiabilidad tiene diversas acepciones, en el contexto del proceso diagnóstico en Psicopatología por fiabilidad se entiende básicamente el *grado de acuerdo* que dos o más jueces logran en la asignación de un individuo a una categoría determinada.

La escasa fiabilidad de los instrumentos diagnósticos ha sido la crítica fundamental a las clasificaciones en Psicopatología (Beck et al., 1962; Kendell et al., 1971; Matarazzo, 1983; Spitzer y Fleiss, 1974). Este panorama desolador ha durado hasta bien entrada la década de los

setenta. Sin embargo, como describiremos más adelante, diversos factores están mejorando esta situación.

Los estudios sobre fiabilidad diagnóstica han demostrado un doble fenómeno. En primer lugar, la mayor fiabilidad se daba en cuadros orgánicos cerebrales (por ejemplo, epilepsia) y drogadicciones, luego en psicosis funcionales y, por último, en neurosis y alteraciones de la personalidad (véase Spitzer y Wilson, 1982). Es decir, la fiabilidad aumenta con la «organicidad» del cuadro. El segundo hallazgo importante en todos estos estudios es que la concordancia diagnóstica disminuye cuanto más se intente especificar el cuadro diagnóstico (Kendell, 1973); es decir, existe un mayor acuerdo en el diagnóstico global de «esquizofrenia» que en el diagnóstico del subtipo de esquizofrenia.

*Estabilidad temporal.* La fiabilidad entendida como estabilidad diagnóstica temporal tiene un valor relativo, pues la condición clínica del paciente puede variar realmente debido a la propia evolución natural del trastorno. Esto es, el estado del paciente puede oscilar de un momento a otro como una característica *intrínseca* a su cuadro (piénsese, por ejemplo, en las variaciones diurnas del estado de ánimo). En consecuencia, una baja estabilidad temporal diagnóstica no significa necesariamente que el diagnóstico psicopatológico sea inadecuado o esté mal efectuado. No obstante, este tipo de índices temporales, empleando seguimientos a corto y largo plazo, pueden ser útiles como unos indicadores auxiliares en estudios de fiabilidad.

Zubin (1967), en una revisión de todos los estudios de fiabilidad publicados hasta el año 1965, demostró que las cifras de estabilidad diagnóstica en una muestra de 11.987 pacientes, presenta-

ban un patrón semejante a las cifras de concordancia diagnóstica. Es decir, la mayor estabilidad temporal del diagnóstico se observaba en los cuadros orgánicos crónicos ( $r_{xx} = 0.90$ ), mientras que las cifras más bajas se daban en los trastornos de personalidad y las neurosis de ansiedad ( $r_{xx} = 0.18$ ).

*Fuentes de variabilidad en el diagnóstico.* Una de las cuestiones básicas para intentar aumentar la fiabilidad de un sistema clasificatorio dado consiste en determinar de dónde proceden las fuentes de error del proceso diagnóstico. En otras palabras, determinar a qué se deben las diferencias habidas entre dos observadores al diagnosticar a un mismo individuo.

En una investigación ya clásica y de gran alcance para la Psicopatología, Wear et al. (1962) demostraron que los desacuerdos diagnósticos interjueces provenían no tanto del hecho de que diferentes observadores empleasen técnicas y procedimientos diferentes de recogida de *datos* (entrevistas, instrumentos psicológicos, etc.) como del empleo de diferentes *criterios* utilizados para establecer el diagnóstico. En efecto, los resultados de dicho estudio pusieron de manifiesto que tan sólo el 5 por 100 de las discrepancias se debían a la técnica de entrevista utilizada, mientras que el 80 por 100 de las discrepancias procedían del empleo de una *nosología vaga e imprecisa*.

Estos datos ponen de manifiesto algo de suma importancia: para conseguir un grado aceptable de acuerdo interjueces, no es tan importante el procedimiento de recogida de datos como los *criterios nosológicos* que se utilizan posteriormente. Todos estos problemas seculares en Psicopatología fueron directamente abordados en los años setenta y han sido

muy atenuados con la implantación del sistema diagnóstico DSM-III (APA, 1980).

Ahora bien, el énfasis en incrementar la fiabilidad es deseable, pero es insuficiente por sí sólo. Se puede, por ejemplo, obtener una excelente fiabilidad evaluando atributos irrelevantes o empleando criterios tan explícitos que no se corresponden apenas a *ningún* caso real (Robins y Helzer, 1986). Por ejemplo, es muy fácil ponerse de acuerdo en si alguien es rubio o no, pero evidentemente este atributo no tiene ninguna *relevancia* psicopatológica.

En línea con lo anterior, es importante advertir que si bien la fiabilidad es una condición previa para la validez (Alvira et al., 1979), no existe una relación lineal entre los índices de fiabilidad y validez. Incluso, es posible que una fiabilidad muy elevada en realidad oculte una baja validez de lo que se está midiendo (recordemos el ejemplo del pelo rubio); de hecho, algunos autores sugieren que, como principio general, se debe sospechar de una alta fiabilidad hasta que no se demuestre que existe una buena validez (Grove, 1987).

### Entrevistas estructuradas

Aunque el empleo de instrumentos diagnósticos no tenga efectivamente un papel tan importante en la mejora de la fiabilidad como el uso de criterios diagnósticos claros y operativos, junto al interés reciente por desarrollar este tipo de criterios, se ha unido el deseo de emplear sistemas de recogida de información que eviten, en la medida de lo posible, la variabilidad interjueces en el tipo de datos recogidos para efectuar el diagnóstico. Uno de estos sistemas es el de las entrevistas estructuradas (EE), es decir,

El empleo sistemático de una serie de preguntas explícitas y ordenadas para explorar la psicopatología del sujeto y, en algunos tipos de entrevistas, conducir a un diagnóstico específico. Aunque como irónicamente señala Zubin (1989), las EE tienen sus antecedentes más remotos en los interrogatorios de la inquisición, su desarrollo científico se ha producido en estos últimos años.

El método básico para la formulación de diagnósticos clínicos es el de la entrevista. Ya en los años cincuenta, Robins y otros investigadores de la Washington University en St. Louis, desarrollaron una serie de EE, no publicadas, para recoger de modo fiable información psicopatológica. Este es el germen que da pie al desarrollo y publicación, en la década de los sesenta, de entrevistas que bien con fines de investigación o fines clínicos sirviesen como procedimientos objetivos de recogida de información psicopatológica.

Una de las ventajas de las EE es que proporcionan un modo homogéneo de interrogación sistemática. Dados los sesgos cognitivos y las rutinas del clínico, el empleo de instrumentos formales de recogida de datos (listas de síntomas, EE, etc.) ha demostrado ser un procedimiento más útil que las entrevistas de formato libre para recoger de modo sistemático, ordenado y eficaz, información del paciente. Incluso un síntoma tan llamativo como las ideas de suicidio, aparece recogido tres veces menos en las historias clínicas efectuadas de modo tradicional, que cuando se emplean instrumentos formales de recogida de datos (Robins y Helzer, 1986).

Esto pone de manifiesto que la fiabilidad del sistema diagnóstico no sólo está estrechamente ligada a las reglas mediante las cuales se efectúa el diagnóstico, sino también al *procedimiento* de re-

cogida de información. Seguramente la empresa más importante para mejorar esta recogida ha sido el desarrollo de entrevistas estructuradas o semiestructuradas que, de modo sistemático, recogen la información necesaria para efectuar un diagnóstico.

Hay mucha diferencia en cuanto al grado de «estructuración» de las entrevistas: desde las que tan sólo plantean preguntas orientativas (por ejemplo, BPRS, Overall y Gorham, 1962) a entrevistas con un formato a modo de árbol de decisión en las que se exige seguir un orden determinado y una codificación rigurosa para después poder efectuar un diagnóstico computarizado (por ejemplo, SADS, Spitzer y Endicott, 1978; PSE, Wing et al., 1974). Además, las EE permiten no sólo cubrir el mismo número de aspectos psicopatológicos relevantes en todos los pacientes examinados, sino que el seguir un guión bien establecido, también permite equilibrar la profundidad en que cada uno de esos aspectos es explorado. A pesar de la formulación completa de preguntas detalladas, los pacientes no se sienten incómodos con el procedimiento y, por el contrario, parece demostrado que el examen minucioso y ordenado de sus problemas puede constituir para ellos un incentivo de la entrevista con el clínico (Helzer, 1981).

El uso de EE efectivamente contribuye a incrementar la fiabilidad diagnóstica. Por ejemplo, en el Estudio Piloto Internacional de la Esquizofrenia, patrocinado por la OMS (World Health Organization, 1973), efectuado simultáneamente en nueve centros de diferentes países, se obtuvieron tasas bastante altas de acuerdo diagnóstico intracentros. Así, en el diagnóstico de la esquizofrenia, se obtuvo una fiabilidad media interjueces de 0.81 ( $N = 190$ ) empleando una EE (en concreto, la PSE de Wing et al., 1974).

En conjunto, todas estas entrevistas comparten ciertas características (véase un ejemplo en la tabla 1):

1. Especifican los *síntomas* a investigar.
2. Proporcionan *definiciones* del ítem.

3. Especifican *preguntas* para explorar los ítems de interés.

4. Proporcionan instrucciones para evaluar presencia y *gravedad* de los elementos estudiados.

Todos estos elementos convierten a estas entrevistas en instrumentos de pri-

**Pensamiento con contenido inusual:** Pensamiento con contenido inusual, extraño o extravagante. Evalúese el grado de rareza, no el grado de desorganización del habla. Delirios manifiestamente absurdos, claramente falsos, o ideas extravagantes expresadas verbalmente. Inclúyanse inserción del pensamiento, robo y transmisión. Inclúyanse delirios de grandiosidad, somáticos y persecutorios, incluso aunque se hayan evaluado en otro apartado.

NE No evaluado

1	Ausente.	
2	Muy ligero:	Ideas de referencia (la gente le mira o se ríe de él o ella). Ideas de persecución (la gente le maltrata). Creencias inusuales en poderes psíquicos, espíritus u OV-NIS. No fuertemente sostenidas. Algunas dudas.
3	Ligero:	Igual que el 2 con plena convicción, pero no delirante.
4	Moderado:	Delirio presente, pero no fuertemente mantenido (funcionamiento no alterado); o bien, delirio encapsulado con plena convicción (funcionalmente no alterado).
5	Moderadamente grave:	Pleno(s) delirio(s) con algo de preocupación o algunas áreas de funcionamiento alteradas.
6	Grave:	Pleno(s) delirio(s) presente(s) con mucha preocupación o muchas áreas de funcionamiento alteradas.
7	Extremadamente grave:	Pleno(s) delirio(s) presente(s) con casi una preocupación absoluta o la mayoría de las áreas de funcionamiento alteradas.

#### Preguntas recomendadas:

- ¿Han tenido las cosas o los sucesos un significado especial para usted?
- ¿Vio alguna referencia hacia usted en los periódicos o en la TV?
- ¿Tiene algún tipo de relación especial con Dios?
- ¿Cómo se explica usted las cosas que le están pasando? (Especificar.)
- ¿Se ha sentido bajo el control de alguna otra fuerza o persona?

**Tabla 1.** Ítem de la nueva versión de la Escala Breve de Evaluación Psiquiátrica (*Brief Psychiatric Rating Scale*, Lukoff y col., 1986), una entrevista estructurada de 23 ítems ampliamente utilizada para evaluar psicopatología general en pacientes psiquiátricos. Una puntuación de 2-3 puntos en el ítem indica una intensidad no patológica.

vilegio para la investigación diagnóstica y psicopatológica en general. Asimismo, el seguimiento de un procedimiento normalizado de recogida de información facilita mucho el uso de algoritmos computacionales diagnósticos, lo que puede tener una gran importancia en el proceso de validación.

Con todo ello se logra reducir varianza en la recogida e interpretación de información, y cubrir sistemáticamente la psicopatología relevante (Spitzer, 1984). Incluso algunas entrevistas, como el SCID (Spitzer, Williams y Gibbon, 1987) pueden por sí mismas contribuir decisivamente a la mejora de la fiabilidad diagnóstica, pues su administración conduce directamente a la formulación de un diagnóstico conforme a un sistema clasificatorio formal (en este caso, el DSM-III).

En áreas en las que tradicionalmente ha existido una baja fiabilidad diagnóstica, se están haciendo esfuerzos para desarrollar EE que contribuyan al aumento de los índices de fiabilidad. Así, por ejemplo, comenzamos a disponer de EE específicas para los trastornos de personalidad (Stangl et al., 1986) o de ansiedad (DiNardo et al., 1983).

El uso de estas entrevistas puede hacer que nos preguntemos, con cierta ironía, si todavía «son necesarios los clínicos» (Spitzer, 1984). Aunque algunas están diseñadas de tal modo que incluso personas *no expertas* puedan administrarlas, obteniéndose índices de fiabilidad satisfactorios (Helzer et al., 1985), la mayoría de las EE siguen requiriendo efectuar juicios clínicos sobre la presencia o no del síntoma y su gravedad.

Aunque las EE proporcionan de modo adecuado y sistemático datos sobre el paciente, no son tan útiles para apresar aspectos tan importantes como el estilo de interacción personal del paciente, sus

actitudes, la detección de signos no verbales, etc., todos ellos inevitablemente presentes durante una entrevista (Zubin, 1989). Por otro lado, las EE suelen consumir una cantidad de tiempo considerable. La SADS, por ejemplo, dura una media de 2 horas, lo que a veces puede limitar su uso.

Además de sus virtudes como instrumentos diagnósticos en la investigación, las EE son excelentes instrumentos didácticos, al obligar a los estudiantes a efectuar un repaso próximo a la exhaustividad de la psicopatología del paciente (Helzer, 1981; Rubinson y Asnis, 1989).

Aunque los clínicos experimentados pueden creer innecesario seguir los pasos detallados de este tipo de entrevistas, su utilidad como instrumento favorecedor de la precisión diagnóstica está demostrada. Las EE son útiles para evitar errores diagnósticos en trastornos tan «bien conocidos» como las depresiones (Spitzer et al., 1982) o las esquizofrenias (Lipton y Simon, 1985), incluso en el caso de que dichos diagnósticos se hayan efectuado siguiendo una serie de criterios diagnósticos precisos. A pesar de que su uso en la práctica cotidiana no sea grande, las EE expresan claramente la dirección actual de la Psicopatología hacia una progresiva objetivización en el diagnóstico y evaluación.

### Índices y procedimientos de fiabilidad

En la década de los cincuenta, la fiabilidad se evaluaba comparando tasas diagnósticas de diferentes hospitales. Más tarde comenzó a utilizarse el procedimiento de emplear diversos evaluadores para diagnosticar un mismo caso. De modo progresivo se fueron introduciendo técnicas estadísticas más sofisticadas

como, por ejemplo, índices «kappa» cuando intervienen más de dos jueces (Uebersax, 1983).

En los años cincuenta era muy frecuente emplear, como índice de fiabilidad, un simple porcentaje de acuerdos. Sin embargo, éste es un índice que no corrige el papel del azar (Spitzer y Fleiss, 1974). En 1960, Cohen propuso el índice «kappa» (K), y desde entonces ha sido el estadístico más empleado para evaluar el grado de acuerdo interjueces (Sprock y Blashfield, 1984; Shrout et al, 1987). El índice K oscila de -1 (desacuerdo sistemático) a 1 (acuerdo sistemático); un índice igual a cero indica que la concordancia diagnóstica es igual que si el procedimiento diagnóstico se hubiese efectuado al azar (por ejemplo, arrojando monedas al aire).

Un sencillo ejemplo puede ayudar a comprender este procedimiento estadístico. En la tabla 2 se indican los resultados hipotéticos del grado de acuerdo diagnóstico alcanzado por dos clínicos en el diagnóstico de la depresión en una muestra de 100 sujetos. Una lectura superficial de los datos podría hacernos suponer que el acuerdo global es del 75 por 100 (es decir, la suma del porcentaje

de casos coincidentemente diagnosticados como «deprimidos» y del porcentaje de casos diagnosticados como «no deprimidos»).

Sin embargo, en la tabla podemos observar que hay una serie de discordancias que han de tenerse en cuenta. El índice kappa justamente proporciona el grado de acuerdo interjueces tras corregir el grado de azar en este acuerdo y considerar el porcentaje de discordancias, según la siguiente fórmula:

$$K = \frac{P_0 - P_a}{1 - P_a}$$

En esta fórmula,  $p_0$  es el porcentaje empírico de acuerdo mientras que  $P_a$  es el porcentaje de acuerdo esperado si el diagnóstico se hubiese efectuado aleatoriamente. El porcentaje  $P_a$  viene formulado por la relación entre la probabilidad de señalar la presencia de un diagnóstico determinado por 2 evaluadores diferentes ( $p_1$  y  $p_2$ , respectivamente):

$$p_a = p_1 p_2 + (1 - p_1)(1 - p_2)$$

		Clínico 2		
		Depresión presente	Depresión ausente	TOTAL
Clínico 1	Depresión presente	50	10	60
	Depresión ausente	15	25	40
TOTAL		65	35	100

**Tabla 2.** Situación hipotética del diagnóstico de «depresión» efectuado por dos clínicos sobre una misma muestra de sujetos.

En otras palabras,  $P_a$  es igual a la suma del producto de las probabilidades marginales que se muestran en la tabla 2. En nuestro ejemplo, el índice de azar sería de  $(0.60 \times 0.65) + (0.40 \times 0.35) = 0.53$ . Si volvemos a la primera fórmula, se obtiene que el índice  $K$  resultante es de 0.47, lo que es notablemente más bajo que el inicial índice de acuerdo no corregido de 0.75.

Aún no se han resuelto algunos problemas conceptuales importantes del índice  $K$ . Por ejemplo, en este índice se efectúa una corrección del azar cuando, en realidad, las confusiones del clínico raramente se rigen por la aleatoriedad. Es decir, cuando existe un desacuerdo entre dos o más clínicos, es muy probable que su desacuerdo se deba a la elección de categorías *limitrofes* y no a la elección de *cualquier* categoría (Sprock y Blashfield, 1984).

En este mismo sentido, sería deseable que en estudios diagnósticos se informase con detalle sobre el *tipo* de errores cometidos en la asignación de categorías a los sujetos. Asimismo, cabría la posibilidad de evaluar los desacuerdos por su magnitud y/o importancia. El «índice Kappa ponderado» de Cohen (1968) podría ser útil para efectuar este análisis (Mezzich, Mezzich y Coffman, 1985). Esta serie de estrategias, en definitiva, podría dar indicios valiosos sobre las fuentes de los errores diagnósticos.

Merecería un comentario más extenso del que nos permiten las limitaciones de este capítulo la metodología de los estudios de fiabilidad. Existen diversos procedimientos (cintas de audio, vídeos, valoración simultánea «in vivo», etc.) para establecer la fiabilidad de un sistema clasificatorio y cada uno de ellos presenta ventajas y dificultades específicas (Grove et al., 1981). Cada uno de estos procedimientos para obtener la fiabilidad inter-

jueces puede, por sí mismo, afectar el grado de acuerdo. Por ejemplo, utilizar cintas de vídeo, tiene la ventaja de que no existen sesgos debidos a la interacción paciente-clínico, pero, como contrapartida, proporciona menos información sobre el paciente (Zubin y Gurland, 1977).

Una observación importante respecto a los índices de fiabilidad es que *no existen* procedimientos para determinar si un índice dado es alto o bajo (Grove et al., 1981). Aunque es habitual señalar límites «acceptables» en torno a 0.5 ó 0.6, para poder efectuar una interpretación adecuada de estas cifras es fundamental tener en cuenta los índices de sensibilidad y especificidad obtenidos como más adelante examinaremos.

Existen, por último, diversos factores que pueden afectar la «fiabilidad del índice de fiabilidad». Así, por ejemplo, en muchos estudios de investigación diagnóstica se le pide al clínico que decida entre un cuadro u otro. Por ejemplo, en la tabla 2 hemos expuesto una típica situación en la que se pide al clínico decidir una de dos opciones dicotómicas. Sin embargo, en situaciones clínicas reales, suelen existir más alternativas diagnósticas (Zwick, 1983). Este tipo de consideraciones tendrán que abordarse en las investigaciones futuras sobre el diagnóstico.

## VALIDEZ

La validez de un sistema diagnóstico señala la precisión del mismo para efectuar una distinción determinada. Normalmente los estudios de validez diagnóstica se centran en dos tipos diferentes de validez: *descriptiva* y *predictiva*. La validez descriptiva denota el grado de homogeneidad que poseen los miembros

adsritos a una misma categoría. A esta homogeneidad pueden contribuir elementos tan diversos como los síntomas, la historia del cuadro, o los antecedentes familiares. En otras palabras, puede intervenir toda aquella información que permita discriminar de un modo adecuado entre miembros pertenecientes a diferentes categorías.

La validez predictiva nos permite predecir el curso, resultado y respuesta al tratamiento. Aunque éste es un importante criterio por su potencial utilidad clínica, Sprock y Blashfield (1984) nos recuerdan que «... dentro de la Psicopatología, no se ha hallado que ningún tratamiento posea una acción específica para una única categoría diagnóstica o que sea efectivo para más de dos tercios de los miembros de una categoría» (p. 296).

La distinción entre ambos tipos de validez a veces puede hacerse borrosa; la validez predictiva puede converger con la validez descriptiva en la diferenciación de grupos de sujetos. Así, por ejemplo, la diferente respuesta terapéutica de unos u otros sujetos (validez predictiva) puede proporcionar información sobre la validez descriptiva de la propia categoría.

En todo caso, como hemos venido señalando reiteradamente, aunque las bases teóricas de la clasificación pueden parecer intachables, la justificación de un sistema de clasificación real sólo puede contrastarse de modo empírico. En este sentido, Blashfield y Draguns (1976b) han defendido que la validez —descriptiva y predictiva— de una categoría debería ser el *criterio determinante* para incluir o no un cuadro dado en una nosología.

Por último, hay que destacar que en Psicopatología y Psiquiatría la validez de constructo es con frecuencia ignorada (Zubin, 1979; Hogan y Nicholson, 1988).

Esto es un grave error, pues, en último término, dicha validez recoge la evidencia empírica global que avala la utilidad explicativa de un concepto (sean teorías etiológicas o categorías nosológicas). Preguntas tales como ¿qué es eso que diagnosticamos como «depresión», o «ansiedad»? ¿es un concepto que se sostiene empíricamente, es robusto?, etc. son, en realidad, preguntas referidas a la validez de constructo. Para responder a estas preguntas esenciales, es necesario recurrir a toda la información que podamos recoger y que sustente la viabilidad de tales conceptos. Este proceso es fundamental para el avance teórico de una disciplina (Silva, 1989). Por ejemplo, un concepto tan extendido en el uso cotidiano como el de «neurosis» ha mostrado tener un escaso apoyo empírico y conceptual y, en las clasificaciones más actuales, finalmente se ha desechado.

## Criterios de validación

La validez global de un sistema clasificatorio ha de someterse a un programa serial de investigación (Skinner, 1981; Feighner, 1987; Zwick, 1983) en el que se tengan en cuenta tanto elementos descriptivos como predictivos. Un modelo acabado de validación debería incluir los siguientes pasos:

1. *Descripción clínica.* Quizás la propiedad más obvia de un sistema diagnóstico sea la de permitir clasificar de un modo adecuadamente *homogéneo* a los individuos. Este debería permitir efectuar tareas diagnósticas sin solapamientos y con la máxima concreción. La delimitación homogénea del cuadro puede conllevar información tan dispar como datos demográficos, factores psicosociales, evolución del cuadro, etc.

Todos los individuos que reciben un determinado diagnóstico deberían presentar un patrón semejante de síntomas y distinguible de otros cuadros clínicos. El empleo de una categoría diagnóstica nos debería permitir idealmente efectuar una distinción *bimodal* en la población (Kendall, 1975).

2. *Historia natural.* Uno de los típicos criterios validatorios de los sistemas clasificatorios en Psicopatología, es la evolución «natural» del cuadro clínico. De hecho, uno de los elementos determinantes adoptados por Kraepelin (1899) en su pionera clasificación lo constituyó la evolución de los pacientes, pues estableció un criterio estricto de validez diacrónica. Así, por ejemplo, el diagnóstico diferencial de la «demencia precoz» se establecía en base al mal pronóstico del cuadro. De hecho, si el paciente diagnosticado con este cuadro acababa recuperándose, Kraepelin consideraba que el diagnóstico efectuado había sido incorrecto (McGuire, 1973).

El curso es, pues, importante para fijar el diagnóstico. Así, por ejemplo, en la DSM-III se determina un mínimo de 6 meses de síntomas continuos para poder definir la presencia de un trastorno esquizofrénico. Aunque el uso de este criterio apenas introduce variaciones diagnósticas en los sujetos esquizofrénicos (Heizer et al., 1981), su uso concuerda con una tendencia deseablemente conservadora en el diagnóstico psicopatológico.

No obstante, este criterio temporal ha de ser considerado con cautela, pues, en nuestra opinión, el denominado «curso natural» parece corresponderse más con una concepción organicista de las alteraciones mentales. A diferencia de lo que pueda suceder con las enfermedades orgánicas, parece improbable que las alte-

raciones mentales tengan una evolución temporal prefijada, ajena al complejo interjuego «biopsicosocial» en el que interviene cualquier respuesta humana (Engel, 1977).

La información sobre el curso puede ser recogida, por ejemplo, en entrevistas estructuradas e, inquiriendo sobre el origen del trastorno, posibles factores de riesgo, número de recaídas, alteraciones en otros miembros de la familia y otra información adicional sobre el curso de la alteración. Todo ello converge finalmente en la búsqueda de homogeneidad de los individuos que componen el grupo diagnóstico.

3. *Resultado del tratamiento.* Tanto este factor como el anterior son elementos importantes de la validación, pues una validación descriptiva no basta por sí sola para determinar la homogeneidad de un caso.

La respuesta al tratamiento puede ser un buen indicador de validación, pero hay que tener en cuenta, asimismo, que puede convertirse en un artefacto. En efecto, la panoplia terapéutica cambia con cierta rapidez y está afectada por multitud de variables, desde la capacidad del terapeuta y las técnicas que usa, hasta las propias disposiciones del sujeto. No obstante, la eficacia de ciertos fármacos como, por ejemplo, las sales de litio en el caso de los trastornos afectivos bipolares, pueden ayudar a proporcionar información convergente sobre la validez del diagnóstico efectuado.

Aunque esta es una vía prometedora, aún no disponemos de información suficientemente detallada sobre la eficacia diferencial de las terapias. No obstante, estos criterios longitudinales plantean el problema de resolver el significado de la «evolución natural», un concepto de profundas raíces médicas y de más difícil aplicabilidad en Psicopatología.

4. *Estudios familiares.* Los estudios genético-epidemiológicos proporcionan una importante información respecto a la consistencia de un determinado trastorno en miembros familiares. Los datos provinientes de este área pueden llegar a formar parte integral de los criterios diagnósticos.

Existe una diversidad de métodos de estudio que permiten efectuar análisis retrospectivos o prospectivos en este ámbito: árboles genealógicos, estudios de gemelos, de adopción, etc. (Feighner y Herbstein, 1987; Pogue-Geile y Harrow, 1984). Este tipo de investigación contribuye a la validación de categorías diagnósticas y permite, además, efectuar hipótesis de diferenciación etiológica, especialmente en el caso de los estudios de gemelos y adopción.

Aunque es difícil desentrañar con claridad los factores genéticos de los aprendidos en una conducta determinada (por ejemplo, Robins y Helzer, 1986), este tipo de investigaciones tendrán un papel cada vez más relevante en investigaciones de validación futuras.

5. *Factores físicos y neurológicos.* Aunque en Psicopatología se estudian básicamente trastornos de origen no orgánico o «funcionales», técnicas recientes como la Tomografía Axial Computarizada (TAC) o la Resonancia Magnética Nuclear, comienzan a ser empleadas con cierta profusión en cuadros psicopatológicos. Estas técnicas, al no ser intrusivas, permiten en algunos casos detectar las áreas corticales que están en funcionamiento mientras el sujeto efectúa determinadas actividades. Los datos procedentes de estas investigaciones aún no se han incorporado a los sistemas diagnósticos, pero sí se utilizan en los procesos de validación subsecuentes (por ejemplo, Vázquez y Ochoa, 1989).

6. *Estudios de laboratorio y evaluación psicodiagnóstica.* La década de los setenta vio resurgir con auge estudios biológicos en la investigación psicopatológica. Dentro de esta oleada, una de las vías de estudio fundamentales consiste en determinar marcadores biológicos del curso, pronóstico e, incluso, inicio de los trastornos. Autores como Akiskal (1980), por ejemplo, han insistido en la necesidad de emplear técnicas biológicas para validar categorías clínicas en el grupo de los trastornos afectivos.

No obstante, los resultados no han sido muy alentadores hasta ahora (Feighner y Herbstein, 1987). Dada la propia inconsistencia de tales marcadores, la gran variabilidad interna de las personas que forman parte de una categoría y, en definitiva, la escasa *especificidad* de los marcadores, no se ha avanzado todo lo rápido que se suponía.

También es posible, como señalamos anteriormente, que parte de la explicación sobre la inconsistencia de los datos existentes hasta ahora se deba a la fragilidad de los propios criterios diagnósticos que se emplean para formar grupos de sujetos. Un ejemplo de esto último puede ser el trabajo de validación diferencial de Walker (1981), en el que se llegó a la conclusión de que las alteraciones esquizoafectivas probablemente son una categoría muy vaga, pues el empleo de pruebas neuromotoras y atencionales no permitió diferenciar los de los trastornos esquizofrénicos o los trastornos afectivos.

7. *Comparaciones cruzadas.* Las comparaciones trans-nacionales, trans-institucionales e históricas pueden ser de importancia para delimitar la consistencia de una categoría diagnóstica. El estudio multinacional sobre la esquizofrenia patrocinado por la OMS (WHO, 1973) o

describan un excesivo significado diagnóstico a un síntoma presente como «hablar de miradas, ojos, etc.» que, *en realidad*, no está ligado de modo especial a una enfermedad como la «paranoia» (Chapman y Chapman, 1969). Es decir, volviendo a un esquema como el de la tabla 3, se da una gran importancia a la casilla (a) sin tener en cuenta la información del resto de las casillas que, por cierto, es igualmente importante (Vázquez, 1985).

**Poder predictivo.** Otro parámetro que proporciona información sobre la utilidad de un sistema diagnóstico es el denominado *poder predictivo positivo* o *negativo* (por ejemplo, Zimmerman et al., 1986; Williams, 1989). El poder predictivo positivo (PPP) consiste en la proporción de individuos clasificados como casos, que son realmente casos. El poder predictivo negativo (PPN) es la proporción de individuos clasificados como no casos, que realmente no son casos. El PPP, el PPN son muy útiles

para determinar el valor potencial de un test o un sistema diagnóstico (Burke y Regier, 1989). Estos dos índices también han de estudiarse para evaluar la eficiencia de un instrumento diagnóstico. Supongamos que queremos averiguar la utilidad diagnóstica de un nuevo test diseñado para diagnosticar un trastorno determinado. Su grado de eficiencia hemos de compararlo con un criterio como, por ejemplo, el diagnóstico de un clínico experto (tabla 4). En el ejemplo que presentamos en la tabla 4, se puede observar que el nuevo test es capaz de diagnosticar correctamente el 77 por 100 de los casos verdaderos de depresión (según el criterio del experto). Su nivel de especificidad es bajo: más un 71 por 100 de los no depresivos son efectivamente identificados como «no depresivos» por dicha prueba. En otras palabras, un 29 por 100 (100-71 por 100) de las personas que no son realmente depresivas son incorrectamente diagnosticadas con este diagnóstico mediante la administración de nuestro test.

	Test o Instrumento diagnóstico		TOTAL
	Depresión presente	Depresión ausente	
Depresión presente	50	15	65
Depresión ausente	10	25	35
TOTAL	60	40	100

Sensibilidad de la prueba:  $50/65 = 0.77$ .

Especificidad de la prueba:  $25/35 = 0.71$ .

Poder predictivo positivo:  $50/60 = 0.83$ .

Poder predictivo negativo:  $25/40 = 0.62$ .

**Tabla 4.** Análisis hipotético de la utilidad de un instrumento diagnóstico. El instrumento es comparado con un criterio externo de referencia (en este caso, el juicio de un clínico experto).

Por último, el PPP es del 83 por 100, lo que indica que del total de casos así clasificados por el instrumento sólo un 83 por 100 son realmente casos. El test es algo peor para rechazar correctamente los no casos reales (PPN = 62 por 100). No hay que olvidar, sin embargo, que unas cifras de validez pobres puede deberse no tanto a la insuficiencia del test como a la insuficiencia del criterio que hemos elegido (es decir, el juicio de un experto). Más adelante discutiremos este problema esencial de la Psicopatología.

### Problemas de la validación

Una de las cuestiones básicas tradicionales que se suscitan en la discusión de los procesos de validación, es el de las relaciones entre la fiabilidad y validez y, como hemos indicado anteriormente, estas relaciones no son sencillas. Una muestra de esto es el hecho de que la fiabilidad puede ser, en sí misma, una medida de validez: un test-retest a las tres meses, a veces puede emplearse como un índice de validez respecto a la estabilidad o curso de un determinado trastorno.

En la validación de un sistema diagnóstico existe el problema crucial de la adopción del *criterio* o punto de referencia externo con el que contrastar dicho sistema. Esto no suele suponer un problema de envergadura para otras ciencias, pues, en general, existen ciertos elementos validantes ajenos al propio sistema diagnóstico. Así, por ejemplo, una infección se evalúa no sólo por una serie de síntomas clínicos, sino también por criterios incontestables, como las pruebas biológicas de laboratorio. Sin embargo, el Psicopatología aún no disponemos de un sistema clasificatorio que

sea independiente del propio sistema (Widiger y Frances, 1987).

La determinación del *criterio* es conceptualmente difícil y se corre el peligro de incurrir en un proceso tautológico. En efecto, es muy frecuente la práctica de emplear instrumentos diagnósticos o clasificatorios que se intentan validar *conforme* a categorías clínicas establecidas previamente. Si el instrumento, o el test, no permiten detectar diferencias entre los grupos diagnósticos, se supone que es una prueba inválida. Sin embargo, es posible que el problema esté en que las categorías sean en sí mismas inválidas o que permitan un excesivo solapamiento entre sujetos adscritos a diferentes grupos. Un ejemplo de este problema es la observación de Helzer y Coryell (1983) de que la escasa validez del test de la supresión de la dexametasona para diagnosticar la depresión, pudiera deberse, más que a un problema intrínseco del test, a las diferencias en los criterios de diagnóstico de depresión que se utilizan en diversos estudios.

Este problema es de difícil y compleja solución en Psiquiatría y Psicología. Las categorías se validan en función de la información convergente procedente de tests, pruebas de laboratorio, etc. Pero a su vez, la validez discriminativa de estas pruebas se establece en la medida en que sean capaces de detectar diferencias entre grupos diagnósticos *previamente* formados según criterios clínicos. La mejor vía para salir de esta situación circular es el intentar que las categorías diagnósticas sean simultáneamente validadas, utilizando el máximo posible de variables (curso clínico, pruebas de laboratorio, etc.).

El ejemplo propuesto en la tabla 4 es paradigmático de este problema. Podemos intentar determinar la validez diagnóstica de un test psicológico o biológico

co o de un nuevo sistema diagnóstico, comparándolos con un criterio externo de «probada» validez. Ahora bien, el problema radica en la propia validez del criterio (en el ejemplo presentado, el juicio de clínicos expertos). Este ejemplo también pone de manifiesto que los grupos diagnósticos formados según el juicio clínico jamás deben considerarse como un buen criterio o punto de referencia intocable *prima facie*. En definitiva, éste es uno de los fascinantes laberintos en los que el diagnóstico, en general, está inmerso.

La validación empírica de un sistema clasificatorio es una tarea siempre inacabada. Sin embargo, en vez de plegarse a críticas apriorísticas, los sistemas de clasificación deben someterse a procesos empíricos que determinen su calidad y viabilidad (H. Skinner, 1981). Las etiquetas diagnósticas, insistimos de nuevo en ello, han de entenderse conceptualmente como entidades hipotéticas, tentativas, y que en modo alguno son «realidades». Como señalaban Spitzer y col. (1975), «... quizás el verdadero valor de unos criterios diagnósticos queda demostrado no por cuánto tiempo permanezcan inalterados, sino por la rapidez con la que facilitan investigación que dé como resultado la modificación y mejora de esos criterios».

## BIBLIOGRAFIA

- ABRAHAM, I. L.: «Diagnostic discrepancy and clinical inference: A social-cognitive analysis», *Genetic, Social and General Psychology Monographs*, 112, 41-102, 1986.
- AKISKAL, H. S.: «External validating criteria for psychiatric diagnosis: Their application in affective disorders», *Journal of Clinical Psychiatry*, 41, 6-14, 1980.
- ALBEE, G. W.: «Notes toward a position paper opposing psychodiagnosis», en: A. R. Maher (Ed.), *New approaches to personality classification*. Nueva York: Columbia University Press, 1970.
- ALVIRA, F.; AVIA, M. D.; CALVO, R., y MORALES, J. F.: *Los dos métodos de las ciencias sociales*. Madrid: Centro de Investigaciones Sociológicas, 1979.
- AMERICAN PSYCHIATRIC ASSOCIATION: *Diagnostic and Statistical Manual of Mental Disorders (3rd ed.)* Washington, D.C.: American Psychiatric Association. [Trad. esp. en Masson, 1983.]
- AMERICAN PSYCHIATRIC ASSOCIATION: *Diagnostic and Statistical Manual of Mental Disorders (3rd ed. Revised)* Washington, D.C.: American Psychiatric Association. [Trad. esp. en Masson, 1988.]
- ARKES, H. R., y HARKNESS, A. R.: «Impediments to accurate clinical judgment and possible ways to minimize their impact», *Journal of Consulting and Clinical Psychology*, 49, 323-330, 1981.
- BANDURA, A.: «The psychology of chance encounters and life paths», *American Psychologist*, 37, 745-755, 1982.
- BECK, A. T.: «Reliability of psychiatric diagnosis: I. A critique of systematic studies», *American Journal of Psychiatry*, 119, 210-216, 1962.
- BERNER, P.; KATSCHNIG, H., y LENZ, G.: «Poly-diagnostic approach: A method to clarify incongruencies among the classification of the functional psychoses», *The Psychiatric Journal of the University of Ottawa*, 7, 244-248, 1982.
- BLASHFIELD, R. K.: *The classification of psychopathology: Neokraepelinian and quantitative approaches*. Nueva York: Plenum, 1984.
- BLASHFIELD, R. K., y DRAGUNS, J. G.: «Evaluative criteria for psychiatric classification», *Journal of Abnormal Psychology*, 85, 140-150, 1976a.
- BLASHFIELD, R. K., y DRAGUNS, J. G.: «Toward a taxonomy of psychopathology», *British Journal of Psychiatry*, 42, 574-583, 1976b.
- BURKE, J. D., y REGIER, D. A.: «Epidemiología de los trastornos mentales», en: J. Talbott, R. E. Hales, y S. C. Yudofsky (Eds.), *Tratado de Psiquiatría*. Barcelona: Ancora, 1989. (orig. 1988).
- CHAPMAN, L. J., y CHAPMAN, J. P.: «Illusory correlation as an obstacle to the use of valid psychodiagnostic signs», *Journal of Abnormal Psychology*, 74, 271-280, 1969.
- COHEN, J.: «Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit», *Psychological Bulletin*, 70, 213-220, 1968.
- COOPER, J. E.; KENDALL, R. E.; GURLAND, B. J.; SHARPE, L.; COPELAND, J. R., y SIMON, R. J.: *Psychiatric diagnosis in New York and London*. Maudsley Monographs. Londres: Oxford University Press, 1972.
- DI NARDO, P. A.; O'BRIEN, G. T.; BARLOW, D. H.; WADDELL, M. T., y BLANCHARD, E. B.: «Reliability of DSM-III anxiety disorder categories using a new structural interview», *Archives of General Psychiatry*, 40, 1.170-1.178, 1983.
- ENGEL, G. L.: «The need for a new medical model: A challenge for biomedicine», *Science*, 196, 129-135, 1977.
- EYSENCK, H. J.; WAKEFIELD, J. A., y FRIEDMAN, A. F.: «Diagnosis and clinical assessment: The DSM-III», *Annual Review of Psychology*, 34, 167-193, 1983.
- FEIGHNER, J. P., y HERBSTEIN, J.: «Diagnostic validity». En C. G. Last & M. Hersen (Eds.), *Issues in diagnostic research*. Nueva York: Plenum, 1987.
- FRICK, T., y SEMMEL, M.: «Observer agreement and reliability of classroom observational measures», *Review of Educational Research*, 48, 137-184, 1978.
- GROVE, W. M.: «The reliability of psychiatric diagnosis», en: C.G. Last & M. Hersen (Eds.), *Issues in diagnostic research*. Nueva York: Plenum, 1987.
- GROVE, W. M., y ANDREASEN, N. C.: «Multivariate statistical analysis in psychopathology», en: T. Millon y G. Klerman (Eds.), *Contemporary directions in Psychopathology: Towards the DSM-IV*. Nueva York: Guilford, 1986.
- GROVE, W. M.; ANDREASEN, N. C.; MCDONALD-SCOTT, P.; KELLER, M. B., y SHAPIRO, R. W.: «Reliability studies in psychiatric diagnosis: Theory and practice», *Archives of General Psychiatry*, 38, 408-413, 1981.
- GUZE, S. B.: «The future of psychiatry: Medicine or social science?», *Journals of Nervous and Mental Disease*, 165, 225-230, 1977.
- HELZER, J. E.: «The use of a structural interview for routine psychiatric evaluations», *Journal of Nervous & Mental Diseases*, 169, 45-49, 1981.
- HELZER, J. E., y CORYELL, W.: «More on DSM-III: How consistent are precise criteria?», *Biological Psychiatry*, 18, 1.201-1.203, 1983.
- HELZER, J. E.; ROBINS, L. N.; CROUGHAN, J. L., y WELNER, A.: «Renard Diagnostic Interview: Its reliability and procedural validity with physicians and lay interviewers», *Archives of General Psychiatry*, 38, 393-398, 1981.
- HELZER, J. E.; STOLZMAN, R. K.; FARMER, A.; BROCKINGTON, I. F., y WORKS, J.: «Comparison with a DIS/DSM-III based physician reevaluation in St. Louis», en: E. Eaton y L. G. Kessler (Eds.), *Epidemiological field methods in psychiatry: The NIMH Epidemiological Catchment Area Program*. Nueva York: Academic, 1985.
- HOGAN, R., y NICHOLSON, R. A.: «The meaning of personality test scores», *American Psychologist*, 43, 621-626, 1988.
- KANFER, F. H., y SASLOW, G.: «Behavioral analysis: An alternative to diagnostic classification», *Archives of General Psychiatry*, 12, 529-538, 1965.
- KENDELL, R. E.: «Psychiatric diagnosis: A study of how they are made», *British Journal of Psychiatry*, 122, 437-447, 1973.
- KENDELL, R. E.: *The role of diagnosis in psychiatry*. Oxford: Blackwell, 1975.
- KENDELL, R. E.; COOPER, J. E.; GOURLAY, A. J.; COPELAND, J.; SHARPE, L., y GURLAND, B. J.: «Diagnostic criteria of American and British psychiatrists», *Archives of General Psychiatry*, 25, 123-130, 1971.
- LIPTON, A. A., y SIMON, F. S.: «Psychiatric diagnosis in a state hospital: Manhattan State revisited», *Hospital & Community Psychiatry*, 36, 368-373, 1985.

- LUKOFF, D.; LIBERMAN, R. P., y NUECHTERLEIN, K.: «Symptom monitoring in the rehabilitation of schizophrenic patients», *Schizophrenia Bulletin*, 12, 578-603, 1986.
- MAGARO, P. A.: *Cognition in schizophrenia and paranoia: The integration of cognitive processes*. Hillsdale, Nueva Jersey: LEA, 1980.
- MATARAZZO, J. D.: «The reliability of psychiatric and psychological diagnosis», *Clinical Psychology Review*, 3, 103-145, 1983.
- MCGUIRE, R. J.: «Classification and the problem of diagnosis», en: H. J. Eysenck (Ed.), *Handbook of Abnormal Psychology*. San Diego, CA: EDITS, 1973.
- MEEHL, P.: «Diagnostic taxa as open concepts: Metatheoretical and statistical questions about reliability and construct validity in the grand strategy of nosological revision», en: T. Millon y G. Klerman (Eds.), *Contemporary directions in Psychopathology: Towards the DSM-IV*. Nueva York: Guilford, 1986.
- MEZZICH, A. C.; MEZZICH, J., y COFFMAN, G. A.: «Reliability of DSM-III vs. DSM-II in child psychopathology», *Journal of the American Academy of Child Psychiatry*, 24, 273-280, 1985.
- MEZZICH, J.: «Diagnosis and classification», en: S. M. Turner y M. Hersen (Eds.), *Adult Psychopathology and diagnosis*. Nueva York: Wiley, 1984a.
- MEZZICH, J.: «Multiaxial diagnostic systems in psychiatry», en: H. I. Kaplan y B. J. Sadock (Eds.), *Comprehensive textbook of psychiatry* (4.ª ed.). Baltimore: Williams y Wilkins, 1984b.
- MEZZICH, J. E.; GOODPASTOR, W., y MEZZICH, A. C.: «Structural issues in diagnosis», en: C. G. Last y M. Hersen (Eds.), *Issues in diagnostic research*. Nueva York: Plenum, 1987.
- MILLON, T.: «The DSM-III: An insider's perspective», *American Psychologist*, 38, 804-814, 1983.
- MILLON, T.: «On the nature of taxonomy in psychopathology», en: C. G. Last y M. Hersen (Eds.), *Issues in diagnostic research*. Nueva York: Plenum, 1987.
- OVERALL, J. E., y GORHAM, D. R.: «The Brief Psychiatric Rating Scale», *Psychological Reports*, 10, 799-812, 1962.
- OVERALL, J. E., y HOLLISTER, L. E.: «Comparative evaluation of research diagnostic criteria for schizophrenia», *Archives of General Psychiatry*, 36, 1.198-1.205, 1979.
- POGUE-GEILE, M. F., y HARROW, M.: «Strategies for Psychopathological Research», en: A. S. Bellack y M. Hersen (Eds.), *Research methods in clinical psychology*, Nueva York: Pergamon, 1984.
- ROBINS, L. N., y HELZER, J. E.: «Diagnosis and clinical assessment: The current state of psychiatric diagnosis», *Annual Review of Psychology*, 37, 409-432, 1986.
- ROSENHAN, D.: «On being sane in insane places», *Science*, 179, 250-258, 1973.
- RUBISON, E. P., y ASNIS, G. M.: «Use of structured interviews for diagnosis», en: S. Wetzler (Ed.), *Measuring mental illness: Psychometric assessment for clinicians*. Washington, D.C.: American Psychiatric Press, 1989.
- SÁNCHEZ DEL RÍO, C.: «Física a lo grande», *Saber leer*. Agosto-Septiembre, 12, 1988.
- SCHEFF, T.: «The labeling theory paradigm», en: C. Eisdorfer, D. Cohen, A. Kleinman y P. Maxim (Eds.), *Models for clinical psychopathology*. Nueva York: SP Medical & Scientific Books, 1981.
- SHROUT, P. E.; SPITZER, R. L., y FLEISS, J. L.: «Quantification of agreement in psychiatric diagnosis revisited», *Archives of General Psychiatry*, 44, 172-177, 1987.
- SILVA, F.: *Evaluación conductual y criterios psicométricos*. Madrid: Pirámide, 1989.
- SKINNER, H. A.: «Toward the integration of classification theory and methods», *Journal of Abnormal Psychology*, 90, 68-87, 1981.
- SPITZER, R. L.: «Psychiatric diagnosis: Are clinicians still necessary?», en: J. B. Williams y R. L. Spitzer (Eds.), *Psychotherapy research. Where are we and where should we go?* Nueva York: Guilford, 1984.
- SPITZER, R. L., y ENDICOTT, J.: «Medical and mental disorder: Proposed definition and criteria», en: R. L. Spitzer y D. Klein (Eds.), *Critical issues in psychiatric diagnosis*. Nueva York: Raven Press, 1978.
- SPITZER, R. L.; ENDICOTT, J., y ROBINS, E.: *Research Diagnostic Criteria for a selected group of functional disorders*. New York: Biometrics Research, New York State Psychiatric Institute, 1975.
- SPITZER, R. L., y FLEISS, J. L.: «A reanalysis of the reliability of psychiatric diagnosis», *British Journal of Psychiatry*, 125, 341-347, 1974.
- SPITZER, R. L.; SKODOL, A. E., y WILLIAMS, J. B.: «Supervising intake diagnosis: A psychiatric "Rashomon"», *Archives of General Psychiatry*, 39, 1.299-1.305, 1982.
- SPITZER, R. L.; WILLIAMS, J. B., y GIBBON, M.: *Structured Clinical Interview for DSM-III-R-Outpatient version (SCID-OP, 1/4/87)*. Biometrics Research Department. New York Psychiatric Institute, 1987.
- SPITZER, R. L., y WILSON, P. T.: «Clasificación en psiquiatría», en: A. M. Freedman, H. I. Kaplan y B. J. Sadock (Eds.), *Tratado de Psiquiatría (Vol. 1)*. Barcelona: Salvat (Orig., 1980), 1982.
- SPROCK, R. K., y BLASHFIELD: «Classification and nosology», en: M. Hersen, A. E. Kazdin y A. S. Bellack (Eds.), *The clinical psychology handbook*. Nueva York: Pergamon, 1984.
- STANGL, D.; PFOHL, B.; ZIMMERMAN, M.; BOWERS, W., y CORENTHAL, C.: «A structured interview for the DSM-III personality disorders: A preliminary study», *Archives of General Psychiatry*, 42, 591-596, 1986.
- SZASZ, T.: «The myth of mental illness», *American Psychologist*, 15, 113-118, 1960.
- UEBERSAX, J. S.: «A design-independent method for measuring the reliability of psychiatric diagnosis», *Journal of Psychiatric Research*, 4, 335-342, 1983.
- VÁZQUEZ, C.: «Limitaciones y sesgos en el procedimiento de la información: Más allá de la teoría del "hombre como científico"», *Estudios de Psicología*, 23-24, 111-133, 1985.
- VÁZQUEZ, C.: «Sistemas de autorreferencia y esquemas cognitivos», *Revista de Psicología General y Aplicada*, 41, 1.095-1.113, 1986.
- VÁZQUEZ, C.; OCHOA, E. F. L., y FUENTENEBRO, F.: «Cognición y esquizofrenia (I). Percepción y atención», *Monografías de Psiquiatría*, 1, 1989.
- WALKER, E.: «Attentional and neuromotor functions of schizophrenics, schizoaffectives and patients with other affective disorders», *Archives of General Psychiatry*, 38, 1.355-1.358, 1981.
- WARD, C. H.; BECK, A. T.; MENDELSON, M.; MOCK, J. E., y ERBAUGH, J. K.: «The psychiatric nomenclature: Reasons for diagnostic disagreement», *Archives of General Psychiatry*, 7, 198-205, 1962.
- WIDIGER, T. A., y FRANCES, A.: «Definitions and diagnoses: A brief response to Morey and McNamara», *Journal of Abnormal Psychology*, 96, 3, 286-287, 1987.
- WILLIAMS, J. B. W.: «Clasificación psiquiátrica», en: J. A. Talbot, R. E. Hales y S. C. Yudofsky (Eds.), *Tratado de Psiquiatría*. Barcelona: Ancora (Original: 1988), 1989.
- WING, J. K., y STURT, E.: *The PSE-ID CATEGO systems. A supplementary manual*. London: Institute of Psychiatry, 1978.
- WING, J. K.; COOPER, J. E., y SARTORIUS, N.: *The measurement and classification of psychiatric symptoms*. Cambridge: Cambridge University Press, 1974.
- WORLD HEALTH ORGANIZATION: *International Pilot Study of Schizophrenia. Vol 1*. Geneve: Author, 1973.
- ZIMMERMAN, M.; CORYELL, W.; CORENTHAL, M. A., y WILSON, S.: «A self-report scale to diagnose major depressive disorder», *Archives of General Psychiatry*, 43, 1.076-1.081, 1986.
- ZUBIN, J.: «Classification of behavior disorders», *Annual Review of Psychology*, 18, 373-406, 1967.
- ZUBIN, J.: «Use of research instruments in psychopathological assessment: Some historical perspectives», en: S. Wetzler (Ed.), *Measuring mental illness: Psychometric assessment for clinicians*. Washington, D.C.: American Psychiatric Press, 1989.
- ZUBIN, J., y GURLAND, B. J.: «The United States-United Kingdom project on diagnosis of the mental disorders», en: L. L.

Adler (Ed.), *Issues in cross-cultural research*. New York: New York Academy of Science, 1977.

ZWICK, R.: «Assessing the psychometric pro-

perties of psychodiagnostic systems: How do the Research Diagnostic Criteria measure up?», *Journal of Consulting and Clinical Psychology*, 51, 117-131, 1983.