

# Instrumentos de evaluación y diagnóstico en drogodependencias

Francisco González Saiz

Centro Provincial de Drogodependencias de Cádiz. Grupo de Investigación en Medicina Psicosocial de la Universidad de Cádiz.

## 1. NECESIDAD E IMPORTANCIA DE LOS INSTRUMENTOS DE EVALUACIÓN EN TOXICOMANÍAS

Puede parecer absurdo, por obvio, señalar la importancia de los instrumentos de evaluación en este área. No obstante, es pertinente ya que su utilización aún es escasa y claramente insuficiente. Esto puede deberse a la limitada disponibilidad en nuestro medio de instrumentos validados, a un conocimiento parcial de los mismos y a una crítica inespecífica sobre su utilidad, unas veces por este desconocimiento y otras por prejuicios de escuela. Los instrumentos de evaluación y diagnóstico en drogodependencias tienen diversas aplicaciones tanto en la clínica como en la evaluación de programas y en la investigación. Su utilidad surge de la necesidad de encontrar herramientas de medida puestas al servicio del método científico para el avance de los conocimientos sobre los trastornos adictivos.

La evaluación de los resultados del tratamiento de los problemas relacionados con la dependencia de sustancias requiere de la existencia de un lenguaje común y de un conjunto de medidas estandarizadas (Rounsaville, 1993). Durante mucho tiempo, y a pesar de las continuas recomendaciones para la utilización de estos instrumentos de evaluación, su uso ha sido escaso hasta hace muy pocos años.

Las ventajas de este tipo de medidas son obvias. Uno de los elementos básicos del método científico es la comprobación sucesiva de hipótesis. Como por lo general, cada estudio de investigación se diseña para responder a una única pregunta, el análisis de la consistencia de los hallazgos debe realizarse a través de la comparación de distintos trabajos. Lógicamente, esto sólo es posible si las unidades de medida y los instrumentos de evaluación son comparables.

Uno de los principales focos de interés de la investigación en el campo de las toxicomanías, es el estudio de la efectividad de las distintas alternativas de tratamiento. La evidencia acumulada hasta el momento actual apoya algunas conclusiones (Anglin y Haser, 1990; Apsler y Harding, 1991; Gerstein, 1992; McLellan *et al.*, 1992; Fuller *et al.*, 1994; *Project MATCH Research Group*, 1997):

1. El tratamiento de la dependencia y de los problemas relacionados con el consumo de sustancias

es más efectivo que el "no tratamiento". Además, el coste económico y social del mismo, se ve compensado por la reducción en el gasto imputable a los problemas que repercuten en otras áreas (p.ej.; disminución ulterior del uso de recursos sanitarios y "casos" evitados).

2. No existe un tratamiento único que sea efectivo para todas las personas con dependencia de sustancias.

3. Todas las modalidades de intervención (programas de mantenimiento con metadona, comunidades terapéuticas, etc.) producen, en la mayoría de los pacientes, una reducción en el consumo de drogas y en la actividad delictiva, mejorando el estado general de salud y la calidad de vida.

4. Las tasas de abandono son relativamente altas en la mayoría de las modalidades de tratamiento, salvo en los programas de mantenimiento con metadona. Esta última es, además, la modalidad más coste-efectiva.

5. En todas las alternativas, los *resultados* de los programas pueden variar, sobre todo, en función de aspectos relativos al *proceso* (p.ej.; "dosis" de actividad asistencial y tiempo de permanencia) y en menor medida de las características de los pacientes al inicio del tratamiento (*estructura*).

Sin embargo, hay todavía muchos aspectos que desconocemos, entre ellos, los siguientes (McLellan y Alterman, 1991; McLellan *et al.*, 1992; Harrison, 1992; *Project MATCH Research Group*, 1997):

1. Aún no se ha establecido con rigurosidad metodológica una tipología multidimensional de pacientes drogodependientes, que pueda ser interpretada desde una perspectiva transteórica.

2. No sabemos si diferentes tipos de tratamientos son más efectivos para determinados grupos de pacientes.

3. También desconocemos el efecto que una combinación particular de programas o una secuencia concreta de componentes pueda tener sobre un tipo específico de usuarios en cada momento de su historia de consumo.

4. Aún carecemos de datos definitivos sobre si el aumento de la permanencia en tratamiento o de la intensidad del mismo mejora los resultados en todos los casos. Se desconoce en qué pacientes y en qué circunstancias debe modificarse el tipo de intervención.

5. Se sabe muy poco sobre la influencia del contexto y del entorno del tratamiento sobre sus resultados.

El desarrollo de la investigación sobre estos aspectos se ve limitado por la falta de comparabilidad entre los resultados de distintos estudios. Ello es debido, en parte, a la falta de consenso entre los autores sobre los criterios de *resultado* del tratamiento que deben considerarse y a la utilización de diferentes instrumentos de evaluación (Darke *et al.*, 1991). Aunque los

hallazgos de un determinado trabajo sean sólidos (validez interna), la consistencia de los mismos y la posibilidad de acumular las evidencias se ven reducidas si los autores no unifican los instrumentos de medida. Para que las conclusiones sean verosímiles, el instrumento de recogida de datos debe ser válido y fiable. Esto viene dado por la estandarización en su diseño y por el proceso de validación del mismo.

Las observaciones clínicas en medicina están sometidas a un proceso de variabilidad que le son inherentes (Sackett, Haynes y Tugwell, 1985). Estos autores clasifican las fuentes de variabilidad en tres grandes grupos; variación individual, variación debida al instrumento y variación debida al observador.

Para Spitzer y cols. (1975), la variabilidad diagnóstica en los trastornos psiquiátricos atiende a cinco orígenes: (1) la *variación de criterio* es la que se observa entre clínicos cuando éstos establecen el diagnóstico sobre criterios nosológicos diferentes, (2) la *variación de información* aparece cuando las decisiones diagnósticas se apoyan sobre distintos tipos y cantidad de información, (3) la *variación de observación* resulta de la distinta forma en que los clínicos interpretan y evalúan las respuestas de los sujetos, (4) la *variación de síntomas individuales (variación de "ocasión")* se observa cuando un sujeto manifiesta síntomas o aspectos diferentes de un mismo trastorno en diferentes observaciones, y (5) la *variación de sujeto* ocurre cuando los pacientes manifiestan trastornos diferentes en distintas observaciones.

Teniendo esto en cuenta, la utilización de criterios nosológicos DSM-IV y CIE-10 para el establecimiento del diagnóstico psiquiátrico sólo reduciría una de estas fuentes de variabilidad entre clínicos (variación de criterio). El uso de entrevistas estandarizadas (p.ej.; el sistema SCAN o la CIDI) reduciría, además, la variación de información y la variación de observación.

Sin embargo, a pesar de la utilización de entrevistas estandarizadas y de criterios nosológicos, en el diagnóstico psiquiátrico existirían fuentes de variabilidad no controladas (Rounsaville, 1993). Así, la variación de criterio seguirá estando presente en la medida en que las definiciones y los criterios no estén suficientemente operativizados. La variación de información seguirá dándose si las preguntas están formuladas de diferente manera o si los sujetos responden a una misma pregunta de formas distintas. La variación de observación depende de la experiencia de cada clínico y del grado de entrenamiento y estandarización en el manejo del instrumento de medida. Finalmente, las variaciones *de síntomas individuales* y *de sujeto* son difíciles de minimizar, debido a la complejidad de los fenómenos que observamos.

La variabilidad del diagnóstico psiquiátrico (incluidos los trastornos relacionados con el consumo, como el *abuso* y la *dependencia*) ha sido objeto de estudio por parte de muchos autores. Sin embargo, aún son muy escasos los trabajos que analicen la variabilidad entre las observaciones diagnósticas de los problemas relacionados con el consumo de sustancias.

En este sentido, consideramos lo siguiente; en primer lugar, el concepto bi-axial de la adicción, por el cual se establece que *dependencia* y *problemas relacionados* son dos constructos distintos y relacionados entre sí, está sobradamente validado; por otro lado, existen criterios nosológicos válidos y aceptados para el diagnóstico de dependencia, así como entrevistas

estandarizadas para su establecimiento; sin embargo, y en nuestra opinión, no existen criterios "nosológicos" unánimemente aceptados sobre lo que entendemos por *problemas relacionados con el consumo de drogas*. Aunque existen entrevistas clínicas estandarizadas para la evaluación multidimensional, como el *Addiction Severity Index* y el *Opiate Treatment Index*, que exploran áreas similares, éstos no hacen referencia a un determinado criterio teórico operativizado y común (como sí ocurre, por ejemplo, entre la entrevista SCAN y los criterios DSM-IV y CIE-10).

Por todo ello, se hace necesaria la implantación progresiva de los instrumentos de evaluación en la rutina clínica diaria de los servicios de drogodependencias. Es un error pensar que su aplicación se limita de forma exclusiva a la investigación y a la evaluación de programas. Precisamente, debido a las dificultades metodológicas que suponen los estudios controlados en los que se analizan la efectividad diferencial de distintas alternativas para un determinado perfil de pacientes (estudios de "matching"), los métodos cuasiexperimentales en los que utiliza información clínica individual registrada en régimen de monitorización continua, cobran cada día más importancia.

## 2. CRITERIOS Y PARÁMETROS DESCRIPTIVOS DE CALIDAD DE LOS INSTRUMENTOS DE EVALUACIÓN EN DROGODEPENDENCIAS

### 2.1. *Concepto general de evaluación instrumental*

La evaluación puede definirse como el proceso de aplicación de un método sistematizado para describir fenómenos u objetos. Su grado de sistematización puede ser muy variable, yendo desde la mera asignación de códigos preestablecidos hasta los sistemas de cuantificación mediante algoritmos. Aunque algunos autores diferencian entre medición (proceso de recogida de la información) y evaluación (interpretación de los resultados), se tiende a denominar como "evaluación" a todo el conjunto del proceso.

La evaluación puede ser subjetiva u objetiva. La *evaluación subjetiva* se caracteriza por la descripción de constructos hipotéticos o intangibles (ej. calidad de vida, depresión, abuso de sustancias) en oposición a las entidades tangibles descritas por las ciencias experimentales como el peso o la altura (*evaluación objetiva*). En ciencias de la salud esta diferencia no siempre es diáfana, ya que existe una enorme carga individual en la interpretación de pruebas complementarias complejas (histología, diagnóstico por imagen, neurofisiología, etc.). Ello determina que muchas normas de calidad sean comunes a los instrumentos objetivos y subjetivos. La evaluación subjetiva es menos precisa, y ha sido infravalorada hasta muy recientemente, pero la creciente demanda de parámetros intangibles como la satisfacción, el apoyo, la autonomía, la calidad de vida o el nivel de discapacidad del sujeto, ha determinado que en la actualidad la utilización de estos parámetros sea imprescindible en cualquier área de salud.

La evaluación puede ser descriptiva o cuantitativa. La *evaluación cuantitativa* consiste en la elaboración de reglas para asignar números a un fenómeno dado, con el fin de cuantificar uno o varios atributos del mismo. Las reglas son una serie codificada de procedimientos para la

asignación de números. Al evaluar un fenómeno concreto es importante situarlo dentro de un modelo categorial o dimensional, y en este segundo caso, delimitar su carácter uni o multidimensional. Cuando un fenómeno complejo se considera en un marco multidimensional, deben delimitarse las dimensiones básicas sobre las que centrar la evaluación, ya que su número real es prácticamente inabarcable. Así, la calidad de vida puede considerarse como una categoría dicotómica (ausente/ presente), como un fenómeno unidimensional (de menos a más calidad de vida general), o como un fenómeno multidimensional. En el caso de la calidad de vida relacionada con la salud podemos considerar una serie de dimensiones básicas: funcionamiento general (psíquico o grado de bienestar, físico o grado de autonomía y funcionamiento sociolaboral), síntomas asociados al trastorno y a su tratamiento, síntomas de distrés psíquico y discapacidad. De forma adicional se pueden añadir: dolor, funcionamiento sexual, relaciones con el personal sanitario; y así sucesivamente.

Los instrumentos de evaluación constan de un número variable de ítems. El ítem es la unidad básica de información de un instrumento de evaluación, y suele componerse de una pregunta y de una respuesta que generalmente es cerrada y permite la asignación de un código. El glosario es un listado adicional de notas explicatorias sobre la definición precisa de cada ítem y sobre cómo combinarlos en categorías o dimensiones (Strömberg, 1988).

El estudio de los instrumentos de evaluación en el área de la salud mental y las toxicomanías se fundamenta en la biometría y en la psicometría. Algunos aspectos de la fiabilidad y muchos elementos de la validez (sobre todo de la validez predictiva) proceden de los conocimientos biométricos aplicados en los estudios de pruebas diagnósticas en medicina, así como en la ingeniería de las telecomunicaciones. Por otro lado, los aspectos de la consistencia interna (entendida como coeficiente de fiabilidad) y el análisis de la estructura de un instrumento proceden de la psicometría clásica en sus distintos enfoques (Teoría Clásica de los Test, Teoría de la Generalizabilidad y Teoría de la Respuesta al Ítem).

## *2.2. Criterios de calidad de los instrumentos de evaluación en toxicomanías*

Existen tres parámetros básicos para evaluar la calidad de un instrumento de medida: su consistencia, su fiabilidad y su validez. En las escalas de seguimiento debe añadirse un cuarto: la sensibilidad al cambio. Otros parámetros a considerar son la redundancia, la transferencia cultural y el "coste-utilidad" de utilización de un instrumento determinado para el propósito del estudio. Desafortunadamente, no existe una definición consensuada de éstos términos en epidemiología, su significado difiere del uso común y varía según el área de estudio e incluso entre diversos autores dentro de una misma área. Este problema se hace especialmente patente en el área de los instrumentos de evaluación subjetiva.

### *a. Consistencia interna*

La consistencia analiza la solidez psicométrica de la escala, esto es, el nivel en que sus diferentes ítems están relacionados entre sí y la posibilidad de efectuar sumatorios de los mismos para obtener puntuaciones globales. Algunos autores incluyen a la consistencia interna dentro de la fiabilidad o de la validez. Según Hernández-Aguado y cols. (1990), consistencia es aquella

"propiedad que define el nivel de acuerdo o conformidad de un conjunto de mediciones consigo mismas". Para evitar esta confusión terminológica, nos referimos aquí a la consistencia *interna* de la prueba. Algunos métodos estadísticos como el análisis factorial, aportan datos tanto de la estructura interna de la escala como de su relación con modelos externos.

La *homogeneidad* indica el grado de acuerdo entre los ítems de la escala, lo que determina si éstos pueden acumularse y dar una puntuación global. Se puede obtener mediante el estudio de la correlación de los ítems con el total (partición media, coeficiente alfa de Cronbach), análisis factorial ("coeficiente de adecuación muestral") o mediante los modelos de objetividad estadística de Rasch (1980).

La "partición media" (*split-half*) estima la homogeneidad en función de la correlación entre dos mitades equivalentes de la escala (ej. ítems de primera mitad *versus* ítems de la segunda mitad o ítems pares *versus* ítems impares). El coeficiente alfa de Cronbach indica el grado en que los diferentes ítems presentan una correlación (la consistencia interna es alta por encima de 0.7) (Bech y cols., 1993). Para las escalas de respuestas dicotómicas se utiliza el test de Kuder-Richardson, que no es más que una adaptación del coeficiente alfa de Cronbach. En el Análisis Exploratorio de Datos que utiliza al Análisis de los Componentes Principales como procedimiento de factorización, la homogeneidad de una escala puede ser apoyada por un "coeficiente de adecuación muestral elevado" (*M.S.A. Test*) (por encima de 0,5). No debe confundirse la homogeneidad con la "unidimensionalidad" tal como la define el Análisis Confirmatorio de Datos (modelos LISREL o LISCOMP).

#### *b. Fiabilidad externa*

La fiabilidad nos indica el grado en que los resultados de un test son reproducibles. Esta medida depende de la estabilidad de las medidas del test, a pesar del cambio de diversos parámetros externos (esto es, no inherentes al test). El estudio de la fiabilidad externa informará sobre la capacidad de reproducción de los resultados del test en distintas situaciones. McDowell (1987) ejemplifica la diferencia entre la validez y la fiabilidad con un excelente símil: un tirador tiene que aprender a acertar en el blanco y después a hacerlo de forma consistente. La validez vendría dada por el grado en que el disparo se acerca a la diana, y la fiabilidad por el grado de aproximación entre sí de una serie consecutiva de disparos. Algunos autores prefieren el término *variabilidad* para describir las diferencias entre resultados obtenidos en diferentes condiciones de evaluación (Hernandez-Aguado *et al.*, 1990).

Un estudio sobre la fiabilidad de una prueba diagnóstica debe incluir al menos un análisis del nivel de acuerdo obtenido al ser evaluada la misma muestra en las mismas condiciones por dos evaluadores distintos (*fiabilidad inter-examinadores*). Se ha señalado la importancia de contar con evaluadores de experiencia similar, en cuanto a formación y uso del instrumento de evaluación a analizar. Andersen (1989) señala otros factores, como la actitud ante las escalas de evaluación y el tipo de preferencia terapéutica, distinguiendo entre fiabilidad inter-observadores e intra-observadores.

La estabilidad de la prueba se obtiene al ser evaluada la misma muestra por el mismo

evaluador en dos situaciones diferentes (fiabilidad test-retest o *variabilidad intra-observadores*). En algunos casos, (psiquiatría infantil, minusvalías psíquicas y a veces en toxicomanías), se obtienen los datos a partir de informadores, por lo que es necesario analizar la concordancia entre los datos obtenidos con el test con la misma muestra y con el mismo evaluador, pero recabando los datos de dos informadores distintos (fiabilidad inter-informadores).

El índice estadístico utilizado para evaluar la concordancia depende de las características de las variables a evaluar. La utilización en varios estudios del coeficiente de concordancia de Kendall (Siegel, 1966) es discutible. En el caso de *variables dicotómicas* o binarias, la concordancia item a item se puede analizar mediante el *porcentaje de acuerdo* y el índice *kappa simple o no ponderado* (Kramer y Feinstein, 1981). El coeficiente de concordancia *kappa* nos informa del nivel de acuerdo obtenido, una vez descartado el acuerdo atribuible al azar. Ello lo hace más fiable que el simple porcentaje de acuerdo. Sin embargo, un mismo valor de *kappa* puede resultar de patrones diferentes de respuesta. Por ello es conveniente señalar también la frecuencia de aparición de cada item y el porcentaje de acuerdo (Costello, 1994), así como el intervalo de confianza (IC) de la estimación de estos índices. Feinstein (1985) propone unos criterios para interpretar los valores del coeficiente *kappa*.

En el caso de *variables ordinales*, el análisis de la concordancia item a item se puede efectuar utilizando el *porcentaje de acuerdo ponderado* y el *índice kappa ponderado* (Kp). Estos se consideran más adecuados que sus análogos no ponderados por dar una medida más real del nivel de acuerdo al ponderar el desacuerdo según el número de rangos que separan la puntuación asignada por un evaluador, de la asignada por el otro. Así, el peso asignado puede ser 0 para el completo acuerdo, 1 cuando hay un rango de diferencia, 2 cuando hay 2 rangos, etc. (Kramer y Feinstein, 1981).

El método de análisis de la concordancia de las puntuaciones globales de un test es controvertido. Habitualmente se utilizan los coeficientes de correlación para analizar el grado de acuerdo. Dichos coeficientes no deberían ser utilizados para analizar la concordancia entre dos evaluaciones; la tendencia puede ser perfecta, con un coeficiente de correlación de 1, y las medidas obtenidas en un laboratorio ser exactamente el doble o exactamente 10 unidades más altas que las correspondientes medidas obtenidas, con lo que la concordancia obtenida entre ambos laboratorios sería inexistente (Feinstein, 1985). Para analizar la fiabilidad de medidas cuantitativas continuas debe utilizarse el Coeficiente de Correlación Intraclases (CCI) (Bartko y Carpenter, 1976). En realidad el CCI no es un estimador único, sino una familia de estimadores con aplicaciones diferenciales (Shrout, 1995).

### *c. Validez*

La validez se define por el grado en que el instrumento mide aquello que realmente pretende medir, y se demuestra cuando la medida predice un criterio o se ajusta de forma consistente con una serie de constructos relacionados en una teoría aceptada. La validez indica qué proporción de la información recogida es relevante a la cuestión formulada. Existen múltiples formas de validez, con el agravante de que algunos autores utilizan un mismo término para definir

conceptos diferentes.

### c.1. Validez simple

Refleja lo que los expertos consideran como medidas significativas.

### c.2 Validez de contenido

Define el grado en que el conjunto de items del test representa adecuadamente el dominio evaluado, esto es, el nivel de representatividad de los items del conjunto de componentes de los que se evalúa (Martinez-Arias, 1995). En realidad, este concepto no difiere mucho de la consistencia, por lo que pueden ser considerados sinónimos.

### c.3. Validez de criterio

Hace referencia al grado en que las puntuaciones de un instrumento se corresponden a un criterio externo o un "patrón-oro", y puedan considerarse un sustituto del mismo (Thiemann, Csernansky y Berger, 1987). La evidencia relacionada al criterio suele considerarse desde dos perspectivas distintas según se trate de relaciones concurrentes o predictivas.

#### i. *Validez concurrente*

Esta nos da una medida de la asociación de las puntuaciones globales con otras escalas de referencia, con un propósito y contenido equivalente. Generalmente se limita al estudio de la correlación entre puntuaciones. Este método puede considerarse como una extensión del análisis factorial para dos grupos de variables. En opinión de algunos autores, este tipo de validez es una forma de validez de constructo. De hecho, constituye una de las estrategias utilizadas para evaluar la validez de un test en ausencia de un "patrón-oro".

#### ii. *Validez predictiva*

La validez predictiva de observación se refiere a la probabilidad de la escala de dar un juicio sobre el fenómeno observado. El análisis de Bayes permite conocer la validez predictiva de un test, su utilidad y su comparabilidad, a partir del análisis de la distribución de los "casos" y "no casos" en una población dada y de su relación con los resultados obtenidos en el test estudiado (positivos y negativos). Una tabla de contingencia de 2x2 expresa esta relación en verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Los coeficientes de validez predictiva obtenidos a partir de la tabla de contingencias son: *sensibilidad*, *especificidad*, *valor predictivo positivo* y *valor predictivo negativo*. Otros

parámetros que pueden obtenerse a través de la aplicación del teorema de Bayes son el *porcentaje de mal clasificados*, la *eficiencia* (proporción de casos bien clasificados en relación al total de evaluados), el *sesgo* (cociente entre los evaluados considerados como positivos y negativos) y el *rendimiento* (casos no detectados por el test en relación al total de casos). Estos coeficientes nos permiten ajustar el punto de corte con respecto al objetivo del estudio.

Si se pretende hacer un estudio de muestreo en dos fases, buscaremos aquel punto de corte que nos permita captar el máximo número de casos, aunque entre ellos se incluyan falsos positivos (especificidad aceptable con una sensibilidad óptima). Si por el contrario pretendemos conocer la morbilidad probable en un población a través de la puntuación en el test, seleccionaremos aquel punto de corte que nos permita descartar el mayor número de "no casos", aunque ello determine la pérdida de algunos falsos negativos (sensibilidad aceptable con un especificidad óptima).

De forma alternativa, el punto de corte idóneo de un test se puede calcular a partir del análisis R.O.C. (*Receiver Operating Characteristics*). Esta técnica fue desarrollada en los años 60 para evaluar la capacidad de discriminación de señales de los controladores de radar. Primero se obtiene una representación gráfica de la tasa de verdaderos positivos (sensibilidad) y la tasas de falsos positivos (1-especificidad) para cada punto de corte. El cálculo del área bajo la curva resultante, nos indica la capacidad discriminante del test a través de todo el *continuum* de morbilidad. Cuando la capacidad discriminante es igual a la obtenida aleatoriamente, se obtiene una diagonal cuya área inferior es de 0.5 (sensibilidad igual a la tasa de falsos positivos). Un test ideal produciría un 100% de verdaderos positivos antes de admitir un solo falso negativo, por lo que el área bajo la curva obtenida sería de 1.0. En la práctica, las áreas bajo la curva oscilan entre 0.5 y 1.0, y permiten un representación gráfica de la capacidad discriminante de diferentes test para un misma dimensión, siendo el mejor aquel que se corresponda con una curva más alejada de la diagonal (Cfr. Thompsom, 1989).

#### c.4. Validez de constructo

Esta se calcula cuando se carece de un "patrón-oro" adecuado (Martinez-Arias, 1995). De hecho, los atributos psicológicos y los procesos mentales son parámetros intangibles que no pueden medirse directamente como la altura o el peso, y por tanto, se trata de conceptos hipotéticos o constructos. Sin embargo, se acepta que muchos constructos psiquiátricos están próximos al criterio (p.ej. depresión, anhedonia), mientras

que otros no pueden ser contrastados con un "patrón-oro" (p.ej. calidad de vida, integración social). Cuando la validez de constructo no se refiere a una entidad nosológica sino a un instrumento de evaluación, se plantea en ocasiones el uso de *análisis factorial exploratorio* como un método de validez de constructo, cuando de hecho, se trata de técnicas sofisticadas para el análisis de su estructura interna o consistencia. Existe, sin embargo, un grupo de técnicas de *análisis factorial confirmatorio*, en las que el investigador trata de verificar una hipótesis explícita que actúa como "criterio" externo. Estas figuran en los programas, LISREL y LISCOMP, que les ha dado nombre.

#### d. Otros parámetros de calidad

Además de la confirmación de su consistencia, fiabilidad y validez, es necesario descartar que la escala sea redundante con otros instrumentos, y sensible al cambio, si se va a evaluar la respuesta a tratamiento.

##### d.1. Redundancia externa

Generalmente, sólo se evalúa la redundancia o el solapamiento de los items de una escala en la fase de construcción de la misma (redundancia interna). Sin embargo, es importante conocer también la redundancia de los items y de las puntuaciones globales con otras escalas de contenido similar (redundancia externa). Este parámetro no es equivalente al de *asociación* obtenido en la validez concurrente.

##### d.2. Sensibilidad al cambio

El estudio de sensibilidad al cambio se puede efectuar mediante el estudio de correlación y análisis de componentes principales en la línea base y tras el período de evaluación (p.ej. después del tratamiento) comparando las estructuras factoriales en ambos puntos. Este tipo de parámetro cobra una gran importancia en los instrumentos para la evaluación de los resultados del tratamiento en drogodependencias. Es necesario que el período de evaluación sea relativamente corto (p.ej. el mes previo) de manera que la medida sea indicativa de la situación actual del paciente (p.ej. el *Addiction Severity Index* y el *Opiate Treatment Index*).

#### 2.3. Selección de un instrumento de evaluación

La selección apropiada de instrumentos de evaluación subjetiva es primordial para cualquier investigación clínica en toxicomanías. Lógicamente deberá usarse aquél que más se adecúe a nuestras necesidades. Sin embargo sorprende el escaso número de revisiones

metodológicas que abordan este tema de forma específica. Bech *et al.* (1993), mencionan una serie de aspectos clave a tener en cuenta en un ensayo clínico:

1. Identificación del motivo: ¿Por qué es deseable usar un instrumento?
2. Identificación del problema: ¿Cuál es el foco del estudio?
3. Identificación de la relevancia: ¿Son relevantes las escalas/cuestionarios utilizados para el estudio?
4. Evaluación del coste-utilidad: ¿Qué información aporta el instrumento en relación al coste de su empleo en tiempo y personal?.

En cualquier caso, la situación ideal sería seleccionar un grupo de instrumentos básicos para su utilización clínica (de forma consensuada por todo el equipo de drogodependencias) y de los que a su vez, pueda obtenerse información suficiente para derivar indicadores para la evaluación del programa.

### 3. DESCRIPCIÓN Y CLASIFICACIÓN DE LOS INSTRUMENTOS DE EVALUACIÓN EN DROGODEPENDENCIAS

Los instrumentos de evaluación en drogodependencias son numerosísimos. En los últimos diez años se ha producido un enorme desarrollo en este campo y actualmente estamos en un momento en el que comienza a generalizarse su uso. Resulta difícil encontrar una forma óptima de clasificar sistemáticamente todo este volumen de información. Así, podemos encontrar en la literatura distintas revisiones, todas incompletas, cada una de las cuales ordena estos instrumentos siguiendo un criterio diferente. En este apartado hacemos una descripción general de los distintos criterios que suelen utilizarse en la clasificación de instrumentos de evaluación en toxicomanías y en salud mental. Posteriormente, hemos realizado una clasificación de acuerdo al criterio del *propósito* del instrumento (área que evalúa, objetivo del estudio y aplicación), ya que ésta nos parece la más práctica y la que más nos interesa cuando tenemos que seleccionar alguno de ellos. Debido al carácter de resumen de esta revisión (como introducción a un módulo de formación), sólo se hace referencia a algunos instrumentos para ejemplificar conceptos. En los contenidos del curso se describirán con más extensión.

#### 3.1. *Criterios generales de clasificación*

Bech y cols (1993) proponen una descripción general de estos instrumentos basada en los objetivos y en la composición de la escala:

- 1) *Area de evaluación*: según este criterio, distinguimos las escalas diagnósticas, sintomáticas, de personalidad y escalas para otros propósitos específicos.
- 2) *Tipo de administración*: escalas para el paciente, el médico u otro personal sanitario.
- 3) *Acceso temporal retrospectivo*: marco temporal de la evaluación.
- 4) *Selección de items* : distingue entre escalas de primera generación (basadas en la experiencia clínica), y de segunda generación (derivadas de las anteriores).
- 5) *Número de items de la escala*.
- 6) *Definición de los items individuales*.

En base a las descripciones efectuadas por otros autores, (Cfr. Thompson, 1989; Wittchen y Essau, 1991), se ha modificado la propuesta original de Bech para permitir una descripción más completa de los diferentes instrumentos en función de su complejidad, propósito y de su construcción (Salvador Carulla, 1996). Los cambios en la terminología con respecto a la utilizada por Bech se detallan en cada apartado. Se describen a continuación estos factores, aportando algunos ejemplos.

#### A. COMPLEJIDAD

En el primer grupo pueden situarse los *questionarios descriptivos* (p.ej.: cuestionarios sociodemográficos) y los *inventarios de síntomas* (ej. inventarios de efectos secundarios al LAAM, utilizados en algunos estudios). Estos instrumentos no permiten una cuantificación de sus items y pueden considerarse como meras listas de comprobación o chequeo.

En un segundo nivel se encuentran las *escalas de evaluación*. Como su nombre indica, éstas permiten una escalación acumulativa de sus items, dando puntuaciones globales al final de la evaluación. Se componen de items individuales, cada uno de los cuales describe una característica bien definida del fenómeno evaluado. Su carácter acumulativo las diferencia de los cuestionarios de recogida de datos y de los meros inventarios de síntomas. Muchos de los instrumentos de evaluación en drogodependencias son escalas de evaluación. Baste citar el *Severity of Dependence Scale* (S.D.S.) (Gossop *et al.*, 1995), el *Severity of Alcohol Dependence Questionnaire* (S.A.D.Q.) (Stockwell, Murphy, Hodgson, 1983), el *Leeds Dependence Questionnaire* (L.D.Q.) (Raistrick *et al.*, 1994), etc.

En un tercer nivel se encuentran las *entrevistas estandarizadas*. Estas se clasifican en función de su objetivo (generales o específicas) y según el nivel de capacitación requerido para su administración, que a su vez depende de la *estructuración* en la formulación de las preguntas y la codificación de las respuestas (a mayor estructuración, menor nivel de capacitación requerido para la administración). Las entrevistas estandarizadas pueden acompañarse de un sistema informatizado de corrección que permite la asignación de criterios diagnósticos. Entre ellas están el *Addiction Severity Index* (A.S.I.) (McLellan *et al.*, 1992) y el *Opiate Treatment Index* (O.T.I.)

(Darke *et al.*, 1991).

Los *sistemas de diagnóstico estandarizado* constituyen el cuarto nivel. Estos proporcionan una codificación de entidades nosológicas con una descripción detallada de cada una de ellas a través de un glosario para facilitar el diagnóstico. Los sistemas de diagnóstico se denominan *operativos* cuando proporcionan una serie de reglas para el diagnóstico, basadas en criterios de inclusión (presencia de un número mínimo de características del fenómeno para su diagnóstico), y de exclusión (despistaje de otras características no relacionadas con el fenómeno). Cuando los criterios de exclusión se refieren a la presencia de otras entidades sindrómicas se consideran que el sistema es *jerárquico*, puesto que efectúa una jerarquía de las entidades nosológicas recogidas en el sistema para su diagnóstico diferencial. Si además permite la codificación de varias entidades o aspectos relacionados en diversos ejes, se considera que el diagnóstico es *multiaxial*. Existen dos sistemas de diagnóstico operativo jerárquico y multiaxial en vigencia en el momento actual: el sistema de investigación de la CIE-10 y el DSM-IV. Para algunos autores los sistemas de diagnóstico no deben ser considerados como un instrumento de evaluación. Sin embargo, en su construcción y en su utilización, los sistemas diagnósticos se ajustan a las reglas generales de la evaluación subjetiva estandarizada. El concepto sindrómico de dependencia de sustancias de Edwards y Gross (1976) se introduce en los sistemas nosológicos actuales a partir del DSM-III-R (Rounsaville *et al.*, 1986).

En un sexto nivel podemos situar una serie de instrumentos de reciente diseño, que generalmente se fundamentan en una entrevista estandarizada. A diferencia de éstas, las *baterías compuestas* de evaluación constan de un conjunto de instrumentos diferentes: cuestionario de recogida de datos, escalas de evaluación incorporadas a la batería, entrevista estandarizada para la recogida de síntomas pasados y/o del estado actual, y sistema informático para el diagnóstico múltiple que permite la codificación diagnóstica según sistemas diferentes. Existen en la actualidad dos baterías compuestas que se ajustan a la descripción anterior: la batería SCAN, desarrollada a partir del PSE (Pull y Wittchen, 1991) (Vazquez-Barquero, 1993), y la batería CASH para evaluación de la esquizofrenia y trastornos del estado de ánimo, desarrollada a partir del SANS/SAPS para evaluación de síntomas positivos y negativos en la esquizofrenia, entre otros instrumentos (Andreasen *et al.*, 1992). En el sistema S.C.A.N., las secciones 11 y 12 evalúan específicamente los trastornos por consumo de "alcohol" y "otras drogas" según criterios DSM-III-R y CIE-10 ("abuso", "dependencia" y "consumo perjudicial").

Otros instrumentos informatizados diseñados fundamentalmente para utilizar en grandes bases de datos son los *sistemas clínicos de información* (Mezzich, 1986).

## B. PROPÓSITO

Al diseñar una escala de evaluación es fundamental tener en cuenta el propósito de la misma en sus diferentes aspectos (patología evaluada, población de referencia, período de evaluación, etc.), la composición de sus ítems y la prevención de sesgos potenciales en su cumplimentación. El propósito de una escala va a determinar el contenido de sus ítems y diversos aspectos relacionados con su estructura. Una escala debe limitarse siempre al área para la que ha sido diseñada, a menos que se efectúe una nueva estandarización de la misma. El propósito se relaciona con la dimensión evaluada, la población objeto de estudio, el período de evaluación y el

tipo de cumplimentación.

#### B.1. Area evaluada.

Las diversas escalas psicosociales evalúan un amplio rango de áreas como síntomas (escalas clínicas), personalidad, adaptación social, familiar, sexual, laboral, discapacidad, etc. Bech (1993) efectúa una distinción entre dos tipos de escalas clínicas: de diagnóstico y sintomáticas. Esta distinción es conflictiva, al existir escalas sintomáticas que han sido utilizadas para el diagnóstico tras calcular el punto de corte idóneo a través de un estudio de validez predictiva (ver parámetros de calidad de una escala), y viceversa.

#### B.2. Objetivo del estudio.

Este nos permite diferenciar entre *escalas generales* (p.ej. para evaluación de caso psiquiátrico) y *escalas específicas* (p.ej. para evaluación de depresión). Las escalas específicas pueden tener a su vez diferentes gradaciones. Wittchen y Essau (1991) distinguen entre escalas basadas en un concepto "amplio" o "restrictivo" de trastorno mental. Los instrumentos más restrictivos priman la especificidad sobre la sensibilidad y viceversa (este factor es particularmente importante en el uso de sistemas de diagnóstico estandarizado).

#### B.3. Marco temporal.

En función de la *estabilidad* del fenómeno evaluado podemos diferenciar entre las *escalas de rasgo*, que evalúan fenómenos relativamente estables a lo largo del tiempo (p. ej. test de personalidad, locus de control); y *escalas de estado*, que evalúan la situación actual del sujeto -generalmente en el último mes- (p.ej. "Estado Psicopatológico" del *Addiction Severity Index*), las últimas semanas o la semana anterior, o los tres días anterior a la evaluación (escalas de "aquí y ahora"). El marco temporal debe detallarse en las instrucciones previas a la administración de la escala.

En las escalas de estado, el *período de evaluación* nos permite diferenciar entre *escalas de detección* (p.ej. para identificación de "probable caso" psiquiátrico, como el GHQ-28 o bien el AUDIT para la detección de "probables casos" de abuso o dependencia del alcohol), *escalas de seguimiento no transicionales* y *transicionales*. Las escalas de seguimiento no transicionales evalúan el cambio en función de la diferencia del puntaje entre dos evaluaciones (ej. *Hamilton Depression Scale*), mientras que las transicionales evalúan directamente el grado de mejoría o empeoramiento experimentado por el paciente entre ambas evaluaciones (ej. escala de cambio del *Clinical Global Impression*). En el estudio de una escala de seguimiento es importante conocer su *sensibilidad al cambio*.

#### B.4. Tipo de administración.

Las *escalas autoadministradas* se diseñan para ser cumplimentadas por el propio sujeto o por un informante. En ocasiones se incluyen items para calibrar la calidad de las respuestas en función de la tendencia a la disimulación y la simulación (p.ej. *EPQ* de Eysenck). Bech y cols. (1993) denomina a este grupo de instrumentos "cuestionarios", sin embargo este término es

demasiado amplio.

Las *escalas heteroadministradas* ("escalas de observador" según Bech), son cumplimentadas por un examinador. Los instrumentos de evaluación heteroadministrados requieren diferentes niveles de capacitación profesional para su aplicación clínica (este factor es particularmente importante en el diseño y administración de entrevistas estructuradas). Las escalas heteroadministradas requieren una estandarización previa del examinador a través de un análisis de su acuerdo con un examinador de referencia (ver fiabilidad inter-examinadores). En ellas están el *Addiction Severity Index* (A.S.I.) (McLellan *et al.*, 1992) y el *Opiate Treatment Index* (O.T.I.) (Darke *et al.*, 1991). Se señalan dos tipos de situaciones extremas en la administración de una escala heteroadministrada: *Situación Alfa* (investigador experto que sigue un interrogatorio cerrado y utiliza una escala con pocos ítems, bien definidos, y que incluyen criterios de mejoría de salud); y *Situación Beta* (evaluador inexperto, que realiza una entrevista abierta, y utiliza una escala con muchos ítems mal definidos y sin criterios de mejoría y salud) (Cfr. Bech y cols, 1993). Algunos instrumentos de evaluación clínica son de *tipo mixto*, incluyendo una sección para síntomas referidos y otra para síntomas observados en la entrevista.

## C. CONSTRUCCIÓN DE LAS ESCALAS DE EVALUACIÓN

Como ya se ha indicado, el ítem es la unidad básica de información de un instrumento de evaluación y generalmente consta de una pregunta y de una respuesta cerrada.

### C.1. Número de ítems.

Puede distinguirse entre *escalas unitarias o globales*, compuestas por un solo ítem y *escalas multi-ítem*. Como regla general, se considera que un fenómeno debe ser evaluado con un mínimo de 6 ítems (Bech y cols., 1993). Generalmente las escalas constan entre 10 y 90 ítems. Diversas escalas están disponibles en varias versiones. Así, el *Alcohol Dependence Data* (A.D.D.), una escala autoadministrada para la evaluación del síndrome de dependencia alcohólica que consta de 39 ítems, también está disponible en una versión breve de 15 ítems, el *Short-form Alcohol Dependence Data Questionnaire* (S.A.D.D.) (Raistrick, Dunbar, Davidson, 1983).

### C.2. Contenido de los ítems.

En función del contenido se distingue entre escalas unidimensionales y multidimensionales. En las *escalas unidimensionales*, más del 80% de los ítems evalúan una sola dimensión (Israel, Korazevic y Sartorius, 1983):

- *Dimensión física*: p.ej.; *Cornell Medical Index* de Ware y cols. (1976).
- *Dimensión psíquica*: p.ej.; *Beck Depression Inventory* de Beck y cols. (1961)
- *Dimensión social*: p.ej.; *Social Adjustment Scale* de Weissman (1975).

En las *escalas multidimensionales* los ítems evalúan dos o tres de las dimensiones

señaladas (ej.; *Addiction Severity Index* y *Opiate Treatment Index*). En las escalas heteroadministradas se distingue también entre los ítems referidos por el paciente (subjetivos) y los observados por el evaluador (objetivos).

El *sesgo de ítem* u *orientación* se refiere a la parte del síndrome que aparece mejor reflejada en la escala, y se representa en un porcentaje de la puntuación máxima teórica para cada categoría de síntomas (Thompson, 1989).

### C.3. Definición y ordenación de los ítems.

La definición de cada ítem debe ser exhaustiva y mutuamente excluyente (criterios de Guilford)(Cfr. Bech y cols., 1993). por otro lado, deben tenerse en cuenta una serie de factores tanto al formular las preguntas y las alternativas de respuesta como al ordenar el conjunto de ítems que componen la escala:

a. *Comprensión* : Es necesario adaptar el lenguaje y el tipo de formulación de las preguntas y respuestas al entorno sociocultural del paciente. Así, por ejemplo, la comprensión del uso de análogos lineales tiende a ser mejor en el medio anglosajón que en la Europa meridional, donde la comprensión de análogos numéricos decimales es mayor. Existen diversos índices de evaluación de la comprensibilidad de un texto (ej. índice de Flesch para el idioma inglés). El problema de la comprensión es sumamente importante en la evaluación de pacientes drogodependientes. Por otro lado, la traducción y adaptación de una escala previamente desarrollada en otro idioma y entorno cultural debe seguir una tecnología específica que incluya un proceso de retro-traducción. Recientemente se han aplicado sistemas más complejos como el de traducción conceptual.

b. *Aceptabilidad* : Es fundamental que los ítems sean aceptables para el sujeto evaluado. La *desideratividad social* es un tipo de sesgo que puede alterar la validez de los resultados en las respuestas (Wittchen y Essau, 1991), y que debe tenerse en cuenta al formular las preguntas de determinados ítems. Este sesgo es importante en la evaluación de las conductas de riesgo de contagio del V.I.H. en población drogodependiente, en las que el sujeto tiende a responder aquello que considera como socialmente más aceptable (Jaccard y Wan, 1995). También es necesario limitar el número de ítems para evitar el *cansancio* y favorecer la colaboración del sujeto (este problema es evidente en cuestionarios o en baterías de más de 100 ítems como el MMPI).

#### c. *Prevención de sesgos en la cumplimentación:*

i. La *acquiescencia* (tendencia a responder afirmativamente a la pregunta) determina la necesidad de alternar preguntas formuladas "en negativo". Sin embargo, este tipo de formulación puede disminuir la comprensión del paciente y la fiabilidad de las respuestas (p.ej. ítems del tipo: "No es cierto que Colón descubrió América"= V/F).

ii. El *error de tendencia central* se refiere a la reticencia a responder las alternativas extremas de un ítem, dando preferencia a las centrales. Este problema afecta principalmente a las escalas analógico-verbales de tres o cinco alternativas (ej: Nada,

Algo, Mucho).

iii. El "*sesgo de lateralidad*" se relaciona con la tendencia a responder más a las alternativas situadas a la derecha o a la izquierda, lo que se incrementa cuando uno de los dos extremos contiene siempre las alternativas "deseables", y puede evitarse alternando primero ítems con alternativas positivas a la izquierda y después ítems con alternativas positivas a la derecha.

Cuando se diseña una escala heteroadministrada deben tenerse en cuenta algunos sesgos específicos:

El *efecto halo* se refiere a la tendencia a efectuar un juicio al inicio de la entrevista (ej. diagnóstico heurístico) que condiciona la cumplimentación de los ítems siguientes. Ello puede ocurrir con la "Valoración de la Gravedad del Entrevistador" (V.G.E.) de las escalas del *Addiction Severity Index* si no se sigue correctamente el procedimiento estandarizado. Asimismo este efecto es importante en el diagnóstico de la patología dual en drogodependencias (Buchanan y Carpenter, 1994).

El *error lógico* se produce al juzgar que todos los ítems aparentemente relacionados deben puntuarse de forma similar (así, puede asumirse que un paciente con una puntuación elevada en "ideas suicidas" puntuará también alto en "desesperanza").

El *error de proximidad* conduce a puntuar de forma similar los ítems adyacentes.

La *varianza terminológica* se relaciona con la atribución de un significado diferente a un mismo término. Este problema afecta sobre todo a las escalas clínicas, dada la diferente interpretación de un término según la escuela psicopatológica o los conocimientos de base del evaluador. Este sesgo puede obviarse incluyendo un glosario terminológico anexo a la escala de evaluación.

#### C.4. Selección de los ítems.

Meehl y Golden (1982) señalan una serie de principios o pasos en la construcción de una escala de evaluación de síntomas:

1. Selección de los ítems en función de su relevancia clínica y su validez.
2. Seleccionar los ítems en función de la correlación interna de los ítems cuando se aplican a un grupo mixto de pacientes (que incluye a pacientes con y sin el síntoma evaluado).
3. Selección de ítems con diferente peso jerárquico (que describan los diversos aspectos del fenómeno evaluado), esto es, que no sean redundantes.
4. Ante una igualdad de factores, seleccionar los ítems con mayor potencial de consenso.

5. Comprobar el rendimiento del grupo de items seleccionado en función de diversos criterios externos (edad, sexo, etc.), con el fin de evaluar su extrapolación.
6. Cuando los pasos 3, 4 y 5 no se puedan efectuar, repetir el análisis con items modificados en cuanto a definición o contenido.

Los items tambien pueden seleccionarse en función de su utilidad. Esta se evalúa de acuerdo con tres criterios (Thompson, 1989):

1. *Calibrado*: frecuencia suficiente de respuestas en un item individual como para garantizar su inclusión en la escala. Arbitrariamente puede fijarse en un 10 %.
2. *Monotonidad ascendente*: el item debe mostrar una correlación lineal significativa con la puntuación global (ver homogeneidad).
3. *Baja dispersión*: con respecto a la linea de regresión de la correlación anterior.

#### C.5. Sistema de codificación de respuestas.

##### a. *Escalas categoriales dicotómicas*:

Presentan un sistema de respuesta de dos alternativas: Si/No o Verdadero/Falso (ej: test de personalidad como el MMPI).

##### b. *Escalas analógicas*:

Pueden diferenciarse en función del sistema analógico utilizado para facilitar la respuesta:

E. analógico-lineal: gradación en una línea de 7 a 10 cms. (ej: Escalas para la evaluación del *craving*).

E. analógico-numérica: Gradación similar a la anterior pero con números (de 0 a 7 ó a 10). En las escalas unitarias termométricas los números se colocan en posición vertical. Estas pueden tambien graduarse de 0 a 100 (ej: Escala de Evaluación de Actividad General -EEAG- del DSM-IV). En ocasiones se combinan análogos visuales y numéricos para aumentar la comprensión.

E. Gráficas: Gradación a través de dibujos. Algunos autores consideran a las escalas gráficas como escalas lineares.

E. Analógico-verbal: Gradación en categorías verbales previamente calibradas (ver escalación de Guttman). Generalmente las opciones de respuesta oscilan entre 3 y 7. Likert considera que 5 era el número de alternativas óptimo. Goldberg por su parte, prefiere utilizar cuatro grados de respuesta para evitar el sesgo de tendencia central. Se

considera que por encima de 6 grados el nivel de fiabilidad disminuye significativamente. Las escalas de gravedad usan más grados que las de detección (p.ej. la puntuación V.G.E. del A.S.I. tiene 9, mientras que el GHQ tiene 4). Estas escalas también reciben el nombre de *Likert* en honor a su introductor, hace 60 años (Cfr. Bech y cols., 1986). Sin embargo, también se denomina así un sistema específico de puntuación, por lo que este uso puede prestarse a confusión.

E. Analógico-categoriales: Se consideran dentro de este grupo una serie de escalas que combinan gradación numérica y verbal (ej. la puntuación E.E.A.G. del DSM-IV).

### C.6. Escalación y puntuación de los items.

Como ya se ha señalado, existen diversos métodos para escalar las distintas alternativas de respuesta en una escala así como los items de la misma (Bech *et al.*, 1993). El sistema de puntuación puede variar sustancialmente de una escala a otra, e incluso en una misma escala, cuando se trata de escalas analógico-verbales. Las categorías de cada item en una escala ordinal suelen oscilar entre 2 y 6 (ausente, dudoso, leve, moderado, marcado y severo). La puntuación depende del patrón de asignación numérica que se seleccione.

Las escalas unitarias de gravedad (no transicionales), suelen tener una puntuación máxima de 8 ó 10 cuando son análogos visuales, y de 7 a 10 cuando se trata de análogos verbales u otras formas combinadas. El EEAG puede puntuarse hasta 99, pero en realidad presenta 10 grados de respuesta en decimales. Las escalas globales unitarias de tipo transicional generalmente son de tipo bipolar, permitiendo una puntuación en sentido negativo y positivo (de mayor empeoramiento a mayor mejoría).

Las escalas verbales multi-item permiten diversas asignaciones numéricas. Así el GHQ de Goldberg permite tres asignaciones diferentes: las dos primeras en base al sistema originariamente propuesto por Likert en los años 30, y un tercero propuesto por el propio Goldberg. La "Escala de Valoración por el Entrevistador" del *Addiction Severity Index* puntúa de acuerdo con el sistema de puntuación propuesto por M. Hamilton, que distingue la opción de ausente (0), dudoso (1), y diversos grados de intensidad (del 2 al 4 ó 5).

### 3.2. Clasificación general de los instrumentos de evaluación en drogodependencias en función del objeto.

En una primera aproximación y desde el criterio del objeto y del área de evaluación, podemos clasificar estos instrumentos en dos grandes grupos; aquellos que se utilizan fundamentalmente en la evaluación clínica y aquellos que se emplean en la evaluación de programas. En el primer caso, se trata de instrumentos dirigidos a la valoración de pacientes drogodependientes, abarcan distintas áreas y tienen distintos ámbitos de aplicación. Aunque muchos de éstos también se utilizan en la construcción de indicadores para la evaluación de programas, existen algunos instrumentos específicos para este fin (que se expondrán en el módulo de formación).

Entendemos la *evaluación clínica* como un proceso amplio y global que implica distintas

aproximaciones valorativas. Miller, Westerberg y Waldron (1995) distinguen en el proceso de evaluación seis funciones distintas pero complementarias entre sí: cribado (*screening*), diagnóstico, valoración, motivación, planificación y seguimiento. Cada una de estas funciones puede cubrirse con la utilización de instrumentos de evaluación específicos. Dentro de cada grupo, todos coinciden en el objeto o área de evaluación, aunque pueden diferir en otros aspectos estructurales (p.ej. complejidad, tipo de administración, etc.).

#### a. Instrumentos de cribado (*screening*)

El cribado tiene por objetivo la detección de "probables casos" de abuso o dependencia de sustancias. Suele ser un procedimiento breve y por lo general se efectúa desde los servicios de atención primaria o como parte de programas epidemiológicos (Cooney, Zweben y Fleming, 1995). El modelo más tradicional, denominado "*hallazgo de casos*" o "*cribado para la detección del trastorno*", consiste en valorar a individuos aparentemente sanos para identificar a aquellos con síntomas iniciales o avanzados de dependencia de sustancias.

Existe un gran número de instrumentos de cribado, tanto para población adulta como para adolescentes. Entre los más conocidos están el *CAGE* (Ewing, 1984), el *Cuestionario Breve para Alcohólicos* (CBA o KFA) de Feuerlein (1976), y el *Michigan Alcoholism Screening Test*, MAST (Selzer, 1971). Más recientemente, entre los instrumentos de cribado en alcoholismo destaca el *Alcohol Use Disorders Identification Test* (AUDIT) de Babor y Grant (1989), una escala de 10 ítems recientemente validada en nuestro medio y que ha demostrado su consistencia interna, fiabilidad y validez predictiva respecto de criterios DSM-IV de abuso y dependencia (Martínez Delgado, 1996). El *CAGE-AID* (Brown, 1992) es una adaptación del *CAGE* para el cribado del consumo de otras drogas.

#### b. Instrumentos de diagnóstico

La segunda función de la evaluación clínica sería el *diagnóstico del caso*, entendido como el procedimiento que lleva a determinar la presencia de síntomas y signos de abuso o dependencia según los criterios nosológicos más aceptados (DSM-IV y CIE-10). La asignación de la categoría diagnóstica de "dependencia de sustancias" es un paso necesario dentro del proceso de valoración clínica, pero por sí sólo, aporta pocos datos para organizar el tratamiento. No obstante se olvida con frecuencia la validez del aspecto dimensional de este diagnóstico (como índice de gravedad de la dependencia) y la necesidad de realizarlo como paso previo al establecimiento del diagnóstico de patología dual.

Los instrumentos de referencia obligada para el diagnóstico de dependencia son las entrevistas clínicas estandarizadas construidas a partir de los criterios de los sistemas nosológicos de la Asociación Americana de Psiquiatría (DSM) y de la Organización Mundial de la Salud (CIE). Estas se diferencian por su grado de estructuración. En este apartado hacemos referencia a las secciones de "drogas" y "alcohol" de estos instrumentos, ya que éstas son entrevistas para la evaluación de cualquier trastorno psiquiátrico.

Las *Entrevistas Clínicas Estructuradas* se caracterizan por tener preguntas y respuestas cerradas, lo que permite su utilización por personas no expertas pero entrenadas. Entre ellas están

la *Diagnostic Interview Schedule* (DIS) (Robins *et al.*, 1981), la *Composite International Diagnostic Interview* (CIDI) (Robins *et al.*, 1988) y su módulo de drogas, la *CIDI-SAM* (Cottler *et al.*, 1989).

Las *Entrevistas Clínicas Semiestructuradas* implican un juicio clínico y se requiere un formación en psicopatología, así como un entrenamiento específico para su manejo. Entre ellas están la *Structured Clinical Interview for DSM-III-R* (SCID-R) (Spitzer *et al.*, 1988) y el *Schedules for Clinical Assessment in Neuropsychiatry* (SCAN) (Wing *et al.*, 1990).

Dentro de este grupo podemos incluir una serie de instrumentos que no son estrictamente "diagnósticos" en el sentido que aquí manejamos (aunque a menudo se confunden). Son escalas de evaluación que generalmente miden la *gravedad* de la "dependencia" según constructos más o menos próximos a los criterios nosológicos más aceptados. Siguen un modelo dimensional de evaluación (de menos a más *grave*) frente al modelo categorial que siguen los criterios diagnósticos ("caso" o "no caso" de dependencia). De todas formas esto es relativo, ya que algunas de estas escalas tienen una función de cribado (depende del análisis de validez realizado: concurrente o discriminante, respectivamente) y por otro lado, las secciones de drogas del S.C.A.N. proporcionan el denominado Índice de Definición (valoración de la *gravedad de la dependencia*). Hay un gran número de instrumentos que pertenecen a este grupo. Baste citar el *Severity of Dependence Scale* (S.D.S.) (Gossop *et al.*, 1995) y el *Leeds Dependence Questionnaire* (L.D.Q.) (Raistrick *et al.*, 1994).

### c. Instrumentos para la evaluación de los problemas relacionados con el consumo.

Estos instrumentos pretenden hacer una estimación del grado de afectación que el consumo de sustancias ha producido sobre un conjunto de dimensiones o áreas de funcionamiento del sujeto como el estado general de salud, problemas laborales, familiares, legales, psicológicos, calidad de vida, etc. La mayoría de los instrumentos de evaluación en drogodependencias pertenecen a este grupo.

Dentro de ellos podemos establecer una diferencia entre los instrumentos *multidimensionales* o generales y los *unidimensionales* o específicos. Los primeros tienen cierto nivel de complejidad, ya que suelen componerse de varias subescalas, cada una de las cuales evalúa una dimensión diferente. Los instrumentos unidimensionales suelen ser escalas de evaluación que valoran la gravedad de problemas específicos relacionados con el consumo (familiar, laboral, psicológico, etc., que por razones de extensión no se citan aquí).

Las entrevistas multidimensionales, que pueden ser para adultos o para adolescentes, se clasifican a su vez en función del grado de estructuración. Entre las *semiestructuradas* están los conocidos *Addiction Severity Index* (A.S.I.) (McLellan *et al.*, 1985, 1992) y *Opiate Treatment Index* (O.T.I.) (Darke *et al.*, 1992). El *Individual Assessment Profile* (I.A.P.) (Flynn *et al.*, 1995) sería un ejemplo de entrevista clínica estructurada y multidimensional. Dentro de este grupo podemos incluir a los *instrumentos multidimensionales de recogida de datos*. Suelen ser entrevistas estructuradas que evalúan distintos aspectos y problemas relacionados con el consumo

de drogas agrupados por áreas, pero no están compuestos por escalas de evaluación (es decir, no hay una "puntuación final" para cada dimensión). Entre ellas estaría la *Encuesta sobre Consumidores de Heroína en Tratamiento* (ECHT)(PNSD, 1996).

d. Instrumentos para el diagnóstico de la comorbilidad psiquiátrica (trastorno o patología dual).

Generalmente, las escalas específicas y las secciones correspondientes de los instrumentos multidimensionales que miden el aspecto o el estado "psicológico" relacionado con el consumo de sustancias (p.ej. GHQ-28, SCL-90, etc.) suelen ser escalas dimensionales de afectación global y su propósito no es el establecimiento de categorías diagnósticas comórbidas (patología dual). Como la determinación de ésta tampoco entraría dentro del concepto de "diagnóstico de caso" que aquí manejamos (presencia de criterios de dependencia), entendemos que el diagnóstico de la patología dual y la exploración de otras conductas adictivas (no químicas) deben incluirse como un apartado independiente pero dentro del proceso de evaluación clínica del paciente con un trastorno adictivo.

Dentro de este grupo incluiríamos las entrevistas clínicas estructuradas y semiestructuradas que ya vimos en el apartado "b" (en sus versiones completas). En cualquier caso, para el diagnóstico de patología dual, tan importante es el instrumento como el procedimiento de evaluación (Weiss, Mirin y Griffin, 1992). Rounsaville (1992; 1993) propone tomar en consideración un determinado período de abstinencia antes de conceder validez a los síntomas psicopatológicos observados, para lo que operativiza cuatro "estados en relación con el consumo".

e. Instrumentos para la evaluación de los aspectos motivacionales y disposicionales.

Dentro de este epígrafe englobaríamos a un conjunto de instrumentos para la evaluación de atributos cognitivos y cognitivo-conductuales relacionados con la dependencia. Una gran parte de los instrumentos de medida en drogodependencias pertenece a este grupo. Entre los aspectos a evaluar están los siguientes: estadios de cambio, motivación para el cambio, balance decisional, expectativas de autoeficacia, nivel de competencia situacional ante situaciones de riesgo, análisis de los episodios de recaídas, etc. Baste citar entre ellos el *University of Rhode Island Change Assessment* (URICA) (McConaughy *et al.*, 1989), el *Drug-Taking Confidence Questionnaire* (DTCQ) (Annis y Martin, 1985), el *Situational Confidence Questionnaire* (SCQ) (Annis, 1987), etc.

#### 4. EL *ADDICTION SEVERITY INDEX* (A.S.I.) Y EL *OPIATE TREATMENT INDEX* (O.T.I).

En este módulo de formación se realiza un entrenamiento básico y estandarización en el manejo práctico de dos instrumentos multidimensionales de gran utilidad: el A.S.I y el O.T.I.

El *Addiction Severity Index* (A.S.I.) (McLellan *et al.*, 1985; 1992) es una entrevista clínica semiestructurada para la evaluación de los problemas relacionados con el consumo de drogas. Consta de seis escalas, cuyas puntuaciones totales proporcionan una medida de la gravedad de

cada uno de estos problemas ("Consumo de drogas y alcohol", "Estado médico", "Estado psicopatológico", "Problemas legales", "Familiar-social" y "Laboral"). Cada escala tiene dos tipos de puntuaciones: la "Valoración de la Gravedad por el Entrevistador" (V.G.E.) es un juicio clínico realizado a través de un procedimiento estandarizado y basado en las puntuaciones individuales de determinados ítems. Su valor oscila en un rango de 0-9 y se interpreta como una medida general del nivel de deterioro en un área determinada. La "Puntuación Compuesta" (P.C.) deriva de la suma ponderada de las puntuaciones de algunos ítems, su valor oscila en un rango de 0 a 1 y proporciona una medida de la gravedad de los problemas actuales evaluados por cada escala (30 días previos). El A.S.I. (*Índice de Gravedad de la Adicción*) ha sido recientemente validado en nuestro medio (González-Saiz, 1997).

El *Opiate Treatment Index* (O.T.I.) (Darke *et al.*, 1992) es también una entrevista clínica semiestructurada para la evaluación de los problemas relacionados con el consumo de drogas. Está especialmente diseñada para la evaluación de los resultados del tratamiento de la dependencia de opiáceos. Está formada por seis escalas cada una de las cuales ofrece una única puntuación total que procede de la suma simple de las puntuaciones de cada uno de sus ítems y se interpreta como una medida de la gravedad en cada una de las dimensiones que evalúan ("Consumo de drogas", "Conductas de Riesgo de Contagio del V.I.H.", "Estado general de salud", "Funcionamiento social", "Escala legal" y "Ajuste psicológico"). El O.T.I. (*Indicador del Tratamiento de la Adicción a Opiáceos*) también ha sido validado en nuestro medio (González-Saiz, 1997).

El interés que suscitan los instrumentos multidimensionales como el A.S.I. y el O.T.I. deriva de su utilidad práctica, al ser unos de los instrumentos que cubren más funciones en la valoración clínica y en la evaluación de programas. Las aplicaciones clínicas pueden ser de cinco tipos: *funciones de cribado, diagnósticas, pronósticas, planificación del tratamiento y evaluación clínica (seguimiento)*. La *utilidad diagnóstica* deriva de su capacidad para establecer un perfil del *estado* del paciente en las dimensiones más frecuentemente relacionadas con el consumo de drogas (entendido como nivel de gravedad o deterioro). También tienen una *utilidad de cribado "múltiple"*, en el sentido de que cada una de sus escalas puede detectar diferencialmente un área disfuncionante, susceptible de una mayor evaluación con instrumentos más extensos y específicos. Por otro lado, y en virtud de las evidencias acumuladas por los estudios de seguimiento, las puntuaciones totales de cada una de estas escalas proporcionan una *estimación pronóstica* general de la evolución del caso. También tienen la capacidad de detectar las necesidades de intervención en cada una de las áreas, sobre la que se basa la *planificación del tratamiento*. Finalmente, una de sus aplicaciones más interesantes está en el campo de la *evaluación clínica (seguimiento del caso)*, lo que permite valorar los resultados del tratamiento de los pacientes en distintas dimensiones.

Por otro lado, estos instrumentos tienen una importante aplicación en la evaluación de programas, ya que la información clínica individual, agregada y sintetizada a través de *indicadores de resultado*, constituye la base de los *sistemas de monitorización de resultados* que, a su vez, son el instrumento fundamental puesto al servicio de los *programas de mejora continua de la calidad asistencial*.

Por todo ello, resulta inaplazable la formación de los profesionales en el manejo clínico de estos instrumentos y sobre todo su aplicación sistemática y rutinaria en los servicios de drogodependencias.

## BIBLIOGRAFÍA

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (Washington D.C., American Psychiatric Association).

Andersen, J.; Larsen, J.K.; Schultz, V.; Nielsen, B.M.; Korner, A.; Behnke, K. *et al.* (1989) Psychiatric Rating Scale: Dimension of schizophrenia: reliability and construct validity, *Psychopathology*, 22: 168-176.

Andreasen, N.C.; Flaum, M.; Arndt, S. (1992) The Comprehensive Assessment of Symptoms and History (CASH): An instrument for assessing diagnosis and psychopathology, *Archives of General Psychiatry*, 49: 615-623.

Anglin, M.D.; Hser, Y. (1990) The treatment of drug abuse, En: Toury, M. y Wilson, J.Q. (Eds.) *Drugs and crime* (Chicago, University of Chicago Press).

Annis, H.M. (1987) *Situational Confidence Questionnaire (SCQ-39)*, Addiction Research Foundation, Toronto, Canada.

Annis, H.M.; Martin, G. (1985) *Drug-Taking Confidence Questionnaire*, Addiction Research Foundation, Toronto, Canada.

Apsler, R.; Harding, W.M. (1991) Cost-effectiveness analysis of drug abuse treatment: current status and recommendations for future research, En: *Background papers on drug abuse financing and services research*, NIDA Research Series N°1 (Rockville, National Drug on Drug Abuse).

Babor, T.; Grant, M. (1989) From clinical research to secondary prevention: International

collaboration in the development of the Alcohol Use Disorders Identification Test (AUDIT), *Alcohol Health and Research World*, 13: 371-374.

Bartko, J.J.; Carpenter, W.T. (1976) On the methods and theory of reliability, *The Journal of Nervous and Mental Disease*, 163 (5): 307-317.

Bech, P.; Malt, U.F.; Dencker, S.J.; Ahlfors, U.G.; Elgen, K.; Lewander, T.; Lundell, A.; Simpson, G.M.; Lingjaerde, O. (Eds.) (1993) Scales for assessment of diagnosis and severity of mental disorders, *Acta Psychiatrica Scandinavica*, 87 (Suppl. 372), 1-87.

Buchanan, R.W.; Carpenter, W.T. (1994) Domains of psychopathology: An approach to the reduction of heterogeneity in Schizophrenia, *Journal of Nervous and Mental Disease*, 182: 193-204.

Brown, R.L. (1992) Identification and office management of alcohol and drug disorders, En: Fleming, M.F.; Barry, K.L. (Eds.) *Addictive Disorders* (St. Louis, Mosby Year Book).

Cooney, N.L.; Zweben, A.; Fleming, M.F. (1995) Screening for alcohol problems and at-risk drinking in health-care settings, En: Hester, R.K.; Miller, W.R. (Eds.) *Handbook of Alcoholism Treatment Approaches. Effective Alternatives* (Massachusetts, Allyn & Bacon).

Costello, C.G. (1994) Advantages of the symptom approach to schizophrenia, En: Costello, C.G., *Symptoms of schizophrenia* (New York, John Wiley & Sons).

Cottler, L.B.; Robins, L.N.; Helzer, J.E. (1989) The reliability of the CIDI-SAM: a comprehensive substance abuse interview, *British Journal of Addiction*, 84: 801-814.

Darke, S.; Ward, J.; Hall, W.; Heather, N.; Wodak, A. (1991) *The Opiate Treatment Index (OTI) Manual*, National Drug and Alcohol Research Technical Report, No 11 (Australia: University of New South Wales).

Darke, S.; Hall, W.; Wodak, A.; Heather, N.; Ward, J. (1992a) Development and validation of a multidimensional instrument for assessing outcome of treatment among opiate users: the Opiate Treatment Index, *British Journal of Addiction*, 87: 733-742.

Derogatis, L.; Lipman, R.; Rockels, K. (1974) The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory, *Behavioral Sciences*, 19: 1-16.

Dudley, D. (1976) The Cornell Medical Index as an adjunct to paraprofessional analysis of alcohol addiction, *Journal of Studies on Alcohol*, 37: 97-103.

Drummond, D.C. (1990) The relationship between alcohol dependence and alcohol-related problems in a clinical population, *British Journal of Addiction*, 85: 357-366.

Edwards, G.; Gross, M.M. (1976) Alcohol dependence: Provisional description of a clinical

syndrome, *British Medical Journal*, 1: 1058-1061.

Ewing, J. (1984) Detecting alcoholism: The CAGE questionnaire, *Journal of the American Medical Association*, 252: 1905-1907.

Feinstein, A.R. (1985) *Clinical Epidemiology* (Philadelphia, Saunders).

Flynn, P.M.; Hubbard, R.L.; Luckey, J.W.; Forsyth, B.H.; Smith, T.K.; Phillips, C.D.; Fountain, D.L.; Hoffman, J.A.; Koman, J.J. (1995) Individual Assessment Profile (IAP): Standardizing the assessment of substance abusers, *Journal of Substance Abuse Treatment*, 12(3): 213-221.

Fureman, I.; McLellan, A.T.; Alterman, A. (1994) Training for and maintaining interviewer consistency with the ASI, *Journal of Substance Abuse Treatment*, 11(3): 233-237.

Gerstein, D.R. (1992) The effectiveness of drug treatment, En: O'Brien, C.P.(Ed.) *Addictive States* (New York, Raven Press).

Goldberg, D.P.; Hillier, V.F. (1979) A scaled version of the General Health Questionnaire, *Psychological Medicine*, 9: 139-145.

González-Saiz, F. (1997) *Estandarización de un instrumento de evaluación multidimensional en los trastornos adictivos*. Tesis Doctoral. Universidad de Cádiz.

González-Saiz, F.; Salvador Carulla, L.; Martínez Delgado, J.M.; López Cárdenas, A.; Ruz Franzí, I.; Guerra Díaz, D. (1997) *Indicador del Tratamiento de la Adicción a Opiáceos*, (Cádiz, Servicio de Publicaciones de la Universidad de Cádiz).

Gossop, M.; Darke, S.; Griffiths, P.; Hando, J.; Powis, B.; Hall, W.; Strang, J. (1995) The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamines users, *Addiction*, 90: 607-614.

Guerra Díaz, D. (1992) *Addiction Severity Index (ASI): un Índice de Severidad de la Adicción*, En: Casas, M. (Ed.) *Trastornos psíquicos en las toxicomanías* (Barcelona, Ediciones en Neurociencias).

Harrison, P.A. (1992) *Chemical dependency treatment accountability plan*, Minnesota Department of Human Services Chemical Dependency Program Division (documento interno).

Hernández-Aguado, I.; Porta, M.; Miralles, M.; García-benavides, F.; Bolúmar, F. (1990) La cuantificación de la variabilidad en las observaciones clínicas, *Medicina Clínica*, 95: 424-429.

Israel, L.; Kozarevic, D.; Sartorius, N. (1984) *Source book for the geriatric assessment: I. Evaluation in gerontology* (Basel, World Health Organization, Karger).

Jaccard, J.; Wan, C.K. (1995) A paradigm for studying the accuracy of self-reports of risk

behavior relevant to AIDS: empirical perspective on stability, recall bias, and transitory influences, *Journal of Applied Social Psychology*, 25: 1831-1858.

Kramer, M.S.; Feinstein, A.R. (1981) Clinical biostatistics: LIV. The biostatistics of concordance, *Clinical Pharmacology Therapy*, 29: 111-123.

Martínez Arias, R. (1995) *Psicometría: Teoría de los Test Psicológicos y Educativos*, (Madrid, Síntesis).

Martínez Delgado, J.M. (1996) *Validación de los cuestionarios breves; AUDIT, CAGE y CBA para la detección precoz del Síndrome de Dependencia del Alcohol en atención primaria*. Tesis Doctoral. Universidad de Cádiz.

McConaughy, E.A.; DiClemente, C.C.; Prochaska, J.O.; Velicer, W.F. (1989) Stages of change in psychotherapy: A follow-up report, *Psychotherapy: Theory, Research, and Practice*, 26: 494-503.

McLellan, A.T.; Alterman, A.I. (1991) Patient treatment matching: a conceptual and methodological review with suggestions for future research, En: Pickens, R.W.; Leukefeld, C.G.; Schuster, C.R., *Improving drug abuse treatment*, NIDA Research Monograph Series, N° 106, (Rockville, National Institute on Drug Abuse).

McLellan, A.T.; Kushner, H.; Metzger, D.; Peters, R.; Smith, I.; Grissom, G.; Pettinati, H.; Argeriou, M. (1992) The fifth edition of the Addiction Severity Index, *Journal of Substance Abuse Treatment*, 9: 199-213.

McLellan, A.T.; Luborsky, L.; Cacciola, J.; Griffith, J.; Evans, F.; Barr, H.L.; O'Brien, C.P. (1985) New data from the Addiction Severity Index: Reliability and validity in three centers, *Journal of Nervous and Mental Disease*, 173(7): 412-423.

McLellan, A.T.; O'Brien, C.P.; Metzger, D.; Alterman, A.I.; Cornish, J.; Urschel, H. (1992) How effective is substance abuse treatment-compared to what?, En: O'Brien, C.P. y Jaffe, J.H. (Eds.), *Addictive States* (New York, Raven Press).

McLellan, A.T.; Woody, G.E.; Luborsky, L.; O'Brien, C.P.; Druley, K.A. (1983b) Increased effectiveness of substance abuse treatment: A prospective study of patient-treatment "matching", *Journal of Nervous and Mental Disease*, 171(19): 587-605.

Mezzich, J.E. (1986) *Clinical care and information systems in psychiatry* (Washington, American Psychiatric Press).

Miller, E.R.; Westerberg, V.S.; Waldron, H.B. (1995) Evaluating alcohol problems in adults and adolescents, En: Hester, R.K.; Miller, W.R. (Eds.) *Handbook of Alcoholism Treatment Approaches. Effective Alternatives* (Massachusetts, Allyn & Bacon).

Plan Nacional Sobre Drogas (1996a) *Encuesta sobre Consumidores de Heroína en Tratamiento (E.C.H.T.)* (documento interno no publicado).

Project MATCH Research Group (1997) Matching alcoholism treatments to clients heterogeneity: Project MATH posttreatment drinking outcomes, *Journal of Studies on Alcohol*, 58: 7-29.

Pull, C.B.; Wittchen, H-U. (1991) The CIDI, SCAN, and the IPDE: Structured diagnostic interviews for ICD-10 and DSM-III-R, *European Psychiatry*, 6: 227-285.

Raistrick, D.; Bradshaw, J.; Tober, G.; Weiner, J.; Allison, J.; Healey, C. (1994) Development of the Leeds Dependence Questionnaire (LDQ): a questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package, *Addiction*, 89: 563-572.

Raistrick, D.; Dunbar, G.; Davidson, R. (1983) Development of a questionnaire to measure alcohol dependence, *British Journal of Addiction*, 78: 89-95.

Robins, L.N.; Wing, J.; Wittchen, H-U; Helzer, J.E.; Babor, T.; Burke, J.; Farmer, A.; Jablenski, A.; Pickens, R.; Regier, D.; Sartorius, N.; Towle, L. (1988) The Composite International Diagnostic Interview: an epidemiological instrument suitable for use in conjunction with different diagnostic systems and in different cultures, *Archives of General Psychiatry*, 45: 1069-1077.

Rounsaville, B.J. (1992) Métodos para el diagnóstico de trastornos psiquiátricos en adictos a opiáceos, en: Casas, M. (Eds.), *Trastornos psíquicos en las toxicomanías*, pp. 33-58 (Barcelona, Ediciones en Neurociencias).

Rounsaville, B.J. (1993) Overview: rationale and guidelines for using comparable measures to evaluate substance abusers, en : Rounsaville, B.J.; Tims, F.M.; Horton, A.M.; Sowder, B.J. (Eds.), *Diagnostic Source Book on Drug Abuse Research and Treatment*, pp. 1-10 (Rockville, U.S. Department of Health and Human services; National Institute on Drug Abuse).

Rounsaville, B.J.; Spitzer, R.L.; Williams, J.B.W. (1986) Proposed changes in DSM-III Substance Use Disorders: Description and rationale, *American Journal of Psychiatry*, 143(4): 463-468.

Sackett, D.L.; Haynes, R.B.; Tugwell, P. (1985) *Clinical Epidemiology: a basic science for clinical medicine* (Boston, Little, Brown & Co.).

Salvador-Carulla, L. (1996) Assessment Instruments in Psychiatry: Description and Psychometric Properties, en : Thornicroft, G.; Tansella, M. (Eds.) *Mental Health Outcome Measures*, pp. 189-206. Springer-Verlag.

Selzer, M.D. (1971) The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument, *American Journal of Psychiatry*, 127: 1653-1658.

Shrout, P.E. (1995) Reliability, en *Textbook in Psychiatric Epidemiology*, M.T. Tsuang; M. Tohen; G.E.P. Zahner (New York, John Wiley & Sons).

- Siegel, S. (1966) *Non-parametric statistics for behavioral sciences* (New York, McGraw-Hill).
- Spitzer, R.L.; Endicott, J.; Robins, E. (1975) Clinical criteria for psychiatric diagnosis and the DSM-III, *American Journal of Psychiatry*, 132: 1187-1192.
- Spitzer, R.L.; Williams, J.B.W.; Gibbon, M.; First, M.B. (1990) *User's guide for the structured clinical interview for DSM-III-R* (Washington, American Psychiatric Press, Inc.).
- Stockwell, T.; Murphy, D.; Hodgson, R. (1983) The Severity of Alcohol Dependence Questionnaire: Its use, reliability and validity, *British Journal of Addiction*, 78: 145-155.
- Strömngren, E. (1988) The lexicon and issues in the relation of psychiatry concepts and terms, En; Mezzich, J.E.; von Cranach, M. (Eds.), *International classification in psychiatry*, (Cambridge, Cambridge University Press).
- Thiemann, S.; Csernansky, J.C.; Berger, P. (1987) Rating scales in research: The case of negative symptom, *Psychiatry Research*, 20: 47-55.
- Thompson, C. (1989) *The instruments of psychiatric research* (Chichester, John Wiley & Sons).
- Üstün, T.B.; Wittchen, H-U. (1992) Instruments for the assessment of substance use disorders, *Current Opinion in Psychiatry*, 5: 412-419.
- Vázquez-Barquero, J.L. (1993) *SCAN: Cuestionarios para la evaluación clínica en psiquiatría* (Madrid, Meditor).
- Vázquez-Barquero, J.L.; Gaité, L.; Artal Simon, J.; Arenal, A.; Herrera Castanedo, S.; Díez-Manrique, J.F.; Cuesta Nuñez, M.J.; Higuera, A. (1994) Desarrollo y verificación de la versión española de la entrevista psiquiátrica "Sistema S.C.A.N." ("Cuestionarios para la Evaluación Clínica en Neuropsiquiatría"), *Actas Luso-Españolas de Neurología Psiquiátrica*, 22(3): 109-120.
- Weiss, R.D.; Mirin, S.M.; Griffin, M.L. (1992) Methodological considerations in the diagnosis of coexisting psychiatric disorders in substance abusers, *British Journal of Addiction*, 87: 179-187.
- Weissman, M.M.; Bothwell, S. (1976) Assessment of social adjustment by patient self-report, *Archives of General Psychiatry*, 33: 111-118.
- Wing, J.K.; Babor, T.; Brugha, T.; Burke, J.; Cooper, J.E.; Giel, R.; Jablenski, A.; Regier, D.; Sartorius, N. (1990) SCAN: Schedules for Clinical Assessment in Neuropsychiatry, *Archives of General Psychiatry*, 47: 589-593.
- Wittchen, H-U; Essau, C.A. (1990) Assessment of symptoms and psychosocial disabilities in

primary care, En: Sartorius, N.; Goldberg, D.; Girolamo, G. de; Costa e Silva, J.; Lecrubier, Y.; Wittchen, H-U, *Psychological disorders in general medical settings* (Toronto, Hogrefe & Huber Publ.).