

PARTE I

CONTENIDOS TEÓRICOS

1.- INTRODUCCIÓN

1.1. – INTRODUCCIÓN

A lo largo de su desarrollo histórico, el concepto de Calidad de Vida se ha caracterizado por su continua ampliación. Si inicialmente consistía en el cuidado de la salud personal, pasa luego a convertirse en la preocupación por la salud e higiene públicas, se extiende posteriormente a los derechos humanos laborales y ciudadanos, continúa con la capacidad de acceso a los bienes económicos, y finalmente se convierte en la preocupación por la experiencia del sujeto de su vida social, de su actividad cotidiana y de su propia salud. El concepto se hace así más rico y complejo cada vez.

En 1948 la Organización Mundial de la Salud (OMS) definió la **salud** como “**el completo estado de bienestar físico, mental y social, y no sólo la ausencia de enfermedad**”, con lo cual el término evolucionó, desde una definición conceptual, hasta métodos objetivos, los cuales, mediante cuestionarios o instrumentos generan escalas e índices que permiten medir las dimensiones que conforman el estado de salud.

En los años sesenta y setenta los científicos sociales empezaron a mostrar interés por el término “**Calidad de Vida**” como consecuencia de la percepción de las desigualdades en la distribución de los recursos, el bienestar de la sociedad y la preocupación por el crecimiento de la población en los países subdesarrollados. Este término tiene su gran expansión a lo largo de los años ochenta.

La Calidad de Vida es un concepto multidimensional que incluye estilo de vida, vivienda, satisfacción en la escuela y en el empleo, así como situación económica (VELARDE-JURADO y AVILA-FIGUEROA, 2002). Por lo tanto, consiste en una sensación de bienestar que puede ser experimentada por las personas y que representa la suma de sensaciones subjetivas y personales de “sentirse bien”.

Según SHIN y JOHNSON (1978) la Calidad de Vida puede ser entendida como la posesión de los recursos necesarios para la satisfacción de las necesidades y deseos individuales, la participación en las actividades que permiten el desarrollo personal y la comparación satisfactoria con los demás, aspectos todos que dependen del conocimiento y la experiencia previa del sujeto.

LEVI y ANDERSON (1980) definen la Calidad de Vida como una medida compuesta de bienestar físico, mental y social, tal como lo percibe cada individuo y cada grupo. Estos autores enumeran los siguientes componentes: salud, alimentación, trabajo, vivienda, seguridad social, vestido, ocio y derechos humanos. Se trata, pues, de un concepto global, completo y multidisciplinar.

FELCE y PERRY (1995) entienden el término de Calidad de Vida como la calidad de las condiciones subjetivas de vida de una persona; después fue entendida como *la satisfacción del individuo con sus condiciones de vida*, y más tarde como una *combinación de las condiciones de vida de una persona junto a la satisfacción que el individuo experimenta*, dando lugar a la definición más actual entendida como la *combinación de las condiciones de vida, y la satisfacción personal, ponderadas por la escala de valores, aspiraciones y expectativas personales*.

La Calidad de Vida es una percepción personal de un individuo, de su situación en la vida, dentro del contexto cultural y de valores en que vive y en relación con sus objetivos, expectativas, valores e intereses (Organización Mundial de la Salud, OMS).

Estas definiciones ponen de manifiesto que, aunque la dimensión objetiva es importante para definir la Calidad de Vida de una persona, las percepciones y las expectativas subjetivas de la misma, traducen esta valoración objetiva en la verdadera Calidad de Vida (CV) experimentada.

La Calidad de Vida tiene diversos determinantes: salud, ingresos económicos, libertad, calidad ambiental, etc. Entre los diferentes elementos integrantes del concepto de Calidad de Vida sobresale el de salud, hasta tal punto que en determinados enfoques no sólo es el valor predominante sino también el aglutinante. En la actualidad se utiliza el concepto de **“Calidad de Vida Relacionada con la Salud” (CVRS)**.

Uno de los factores que más influyen en la CVRS son los síntomas. En la Enciclopedia de Bioestadística, se define el *síntoma* como “la percepción del paciente de un estado anormal físico, emocional, o estado cognitivo”. El estado funcional incluye cuatro dimensiones: física, fisiológica, social y la actividad del paciente. En las percepciones del Estado de Salud influyen tanto la evaluación del paciente de su estado de salud

pasada y presente, como sus perspectivas de futuro, y la preocupación por la salud. Todos estos factores influyen en la evaluación de la Calidad de Vida global. (Figura 1.1).



Figura 1.1.- Factores que influyen en la evaluación de la Calidad de Vida global.

A menudo, el término Calidad de Vida se utiliza cuando a un paciente se le informa de un diagnóstico. Esto lleva a confundir el término CVRS con efectos secundarios y/o síntomas, y ambos conceptos no son equivalentes con CVRS, a pesar de que tanto los síntomas como los efectos secundarios influyen en la autopercepción de los individuos a la hora de evaluar su CV.

La Calidad de Vida Relacionada con la Salud o salud percibida integra aquellos aspectos de la vida que están directamente relacionados con el funcionamiento físico y mental y con el estado de bienestar, que pueden ser agrupados en cuatro apartados:

- **Estado físico y capacidad funcional.**
- **Estado psicológico y bienestar.**
- **Interacciones sociales.**
- **Estado económico y sus factores.**

La CVRS constituye una importante variable de medida subjetiva del impacto que la enfermedad y su tratamiento producen en la vida del sujeto. Su valoración nos permite detectar alteraciones e intervenir precozmente, así como establecer comparaciones entre las distintas opciones terapéuticas. La evaluación de la CVRS en medicina nos aporta un resultado final de salud que se centra en la persona, no en la enfermedad; en cómo se siente el paciente, independientemente de los datos clínicos.

Este término ha evolucionado desde una definición conceptual hasta una serie de escalas que permiten medir la percepción general de la salud. Considerando que las expectativas de salud, pueden afectar la Calidad de Vida, dos personas con el mismo estado de salud tienen diferente percepción personal de la salud. En este contexto es el paciente quien debe emitir el juicio perceptivo de Calidad de Vida.

Sin embargo, no existe un consenso sobre la definición de Calidad de Vida, e incluso en la literatura se utilizan los términos “estado de salud”, “estado funcional”, “Calidad de Vida” o “Calidad de Vida Relacionada con la Salud” como sinónimos, cuando realmente son términos relacionados, pero responden a conceptos distintos.

Este término (CVRS) es más restringido, pero aun así puede ser complejo, como se puede ver en las definiciones que de él se hacen:

“Valor asignado a la duración de la vida modificado por la deficiencia, el estado funcional, la percepción de la salud y la oportunidad social debido a una enfermedad, accidente, tratamiento, o política determinada. Este valor es individual y depende de la percepción subjetiva de cada individuo y de su entorno político, social y económico” (PATRICK y ERICKSON, 1988).

Para BOWLING (1994) el concepto de CVRS representa las respuestas individuales a los efectos físicos, mentales y sociales que la enfermedad produce sobre la vida diaria, la cual influye sobre la medida en que se puede conseguir la satisfacción personal con las circunstancias de la vida. Abarca más que un adecuado bienestar físico, incluyendo también las percepciones de éste, un nivel básico de satisfacción y un sentido general de autoestima.

JUNIPER (1997) la define como una componente de la Calidad de Vida global que está determinada primordialmente por el estado de salud de la persona y que puede ser influenciado por intervenciones clínicas.

KAPLAN y BUSH (1982) propusieron el uso del término Calidad de Vida Relacionada con la Salud para centrarse en el papel específico de los efectos de salud sobre las percepciones individuales de bienestar, distinguir estos de la satisfacción profesional y

de los factores ambientales. En la literatura médica, los términos Calidad de Vida y Calidad de Vida Relacionada con la Salud han llegado a ser permutables.

Para la Organización mundial de la Salud, es “un concepto amplio que incorpora, de forma compleja, la salud física de los individuos (síntomas subjetivos como el dolor, la fatiga, dificultades para caminar, etc.) estado psicológico, nivel de independencia, relaciones sociales (limitaciones de los roles, conexiones sociales, etc.), creencias personales, y sus relaciones con características sobresalientes del ambiente (ahorros, fianzas, transporte, educación, etc.)”.

Para nosotros por lo tanto: La Calidad de Vida Relacionada con la Salud agrupa, pues, tanto elementos que forman parte del individuo como aquellos externos que interactúan con él y pueden llegar a cambiar su estado de salud.

Todas estas definiciones hacen hincapié en el carácter subjetivo de la evaluación de CVRS, prestando especial atención a la evaluación individual de cada paciente. Es importante señalar que el “bienestar” o la condición de salud de un individuo no pueden medirse directamente. Sólo podemos hacer inferencias de indicadores de los síntomas y de la información que el sujeto nos da de la percepción de su CV.

Como muchos de los componentes de la Calidad de Vida no pueden ser observados directamente, son comúnmente evaluados de acuerdo a los principios clásicos de la teoría de respuesta al ítem: hay una verdadera Calidad de Vida que no puede ser medida directamente, pero puede ser medida de forma indirecta a partir del análisis de las respuestas de los individuos a un cuestionario que comprende distintas dimensiones (o dominios). Cada dimensión se refiere al área que estamos tratando de medir y se construye a partir de cuestiones conocidas como ítems, cuyos niveles de respuesta están categorizados en escalas; cada uno de los ítems mide el mismo concepto o constructo (en ocasiones se utiliza el término “factor”).

1.2.- MEDIDA DE LA CALIDAD DE VIDA RELACIONADA CON LA SALUD

Hoy en día existe una preocupación creciente por la Calidad de Vida y su medida en el entorno sanitario, debido a que grandes sectores de la población padecen enfermedades crónicas que reducen su Calidad de Vida. Ésta no es la única razón para medir la Calidad de Vida Relacionada con la Salud; tiene el valor añadido de evaluar las intervenciones sanitarias no sólo en la práctica clínica sino desde el punto de vista de la gestión y la política sanitaria. Así, las medidas de Calidad de Vida tiene los siguientes usos potenciales en la práctica rutinaria (HIGGINSON y CARR, 2001):

- Pueden identificar qué problemas son más importantes. Esto resulta especialmente útil cuando los pacientes presentan múltiples problemas (HIGGINSON, 1994). Además, estas medidas pueden utilizarse para captar información que parece no tener ninguna relevancia clínica pero que podría explicar la gravedad de la enfermedad (STOWERS y col., 1999).

- Pueden ayudar a los pacientes a comunicar sus problemas, ya que la medida proporciona información clara sobre una serie de problemas.

- Pueden servir para detectar problemas ocultos. Algunos problemas de los pacientes pueden pasarse por alto a menos que se pregunte específicamente por ellos. Esto ocurre fundamentalmente con los problemas psicológicos y sociales (MAGUIRE y col., 1999).

- Pueden ayudar a identificar el resultado preferido del paciente así como sus objetivos en relación con el tratamiento. Si éstos no se conociesen, el tratamiento podría no cumplir las expectativas del paciente, lo que podría afectar la continuidad del mismo y la satisfacción del paciente con la atención recibida.

- Pueden controlar los cambios o respuestas al tratamiento, complementando las pruebas de laboratorio y las pruebas clínicas.
- Pueden permitir a los clínicos tener una visión general de los problemas de un individuo, ya que comprenden información sobre factores importantes para la persona.
- Pueden asegurar que la intervención sanitaria se concentre en lo que es importante para los pacientes, más que en los aspectos técnicos de la calidad. La revisión de los objetivos del cuidado y la comprobación de su logro puede ayudar al personal sanitario a prepararse para mejorar en el futuro el tratamiento de problemas similares.
- Pueden usarse, junto con la mortalidad y la satisfacción, en la gestión y política sanitarias para evaluar los servicios y los tratamientos (Figura 1.2).

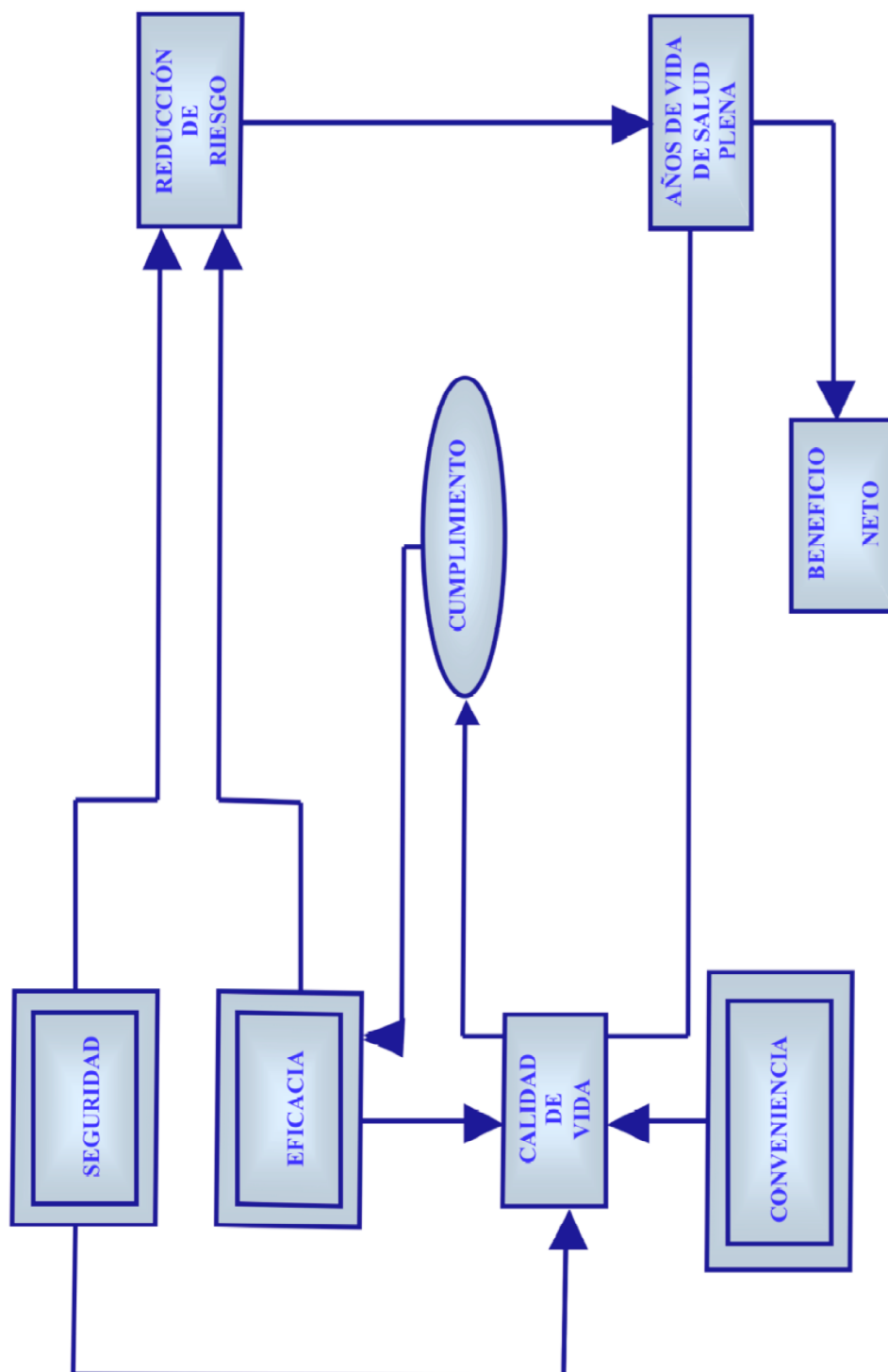


Figura 1.2- PAPEL DE LA CALIDAD DE VIDA EN LA DETERMINACIÓN DEL BENEFICIO NETO DE LA TERAPIA DE UNA ENFERMEDAD CRÓNICA.

TRADUCIDO DE TESTA y col. (1996).

Este hipotético modelo muestra la relación entre las influencias relacionadas con el tratamiento (seguridad, eficacia y conveniencia), las características de los pacientes (cumplimiento), resultados medibles (calidad de vida, reducción del riesgo, años de vida sana y beneficio neto).

1.2.1.- ALGUNOS TÉRMINOS BÁSICOS

En estudios de Calidad de Vida Relacionada con la Salud (CVRS), se hace uso de términos que comenzamos aclarando: medida, valoración, instrumento, test, factor, dimensión, cuestionario, ítem, escala, factor, perfil, índice.

Medir es el proceso de asignar números o categorías a un fenómeno, de acuerdo a unas reglas, de tal forma que las propiedades de los atributos puedan ser representadas por las propiedades de los números, de manera fiable.

Por valoración¹ se entiende un método de medida que incluye un juicio humano. Entendida la Calidad de Vida como una percepción subjetiva del individuo, la evaluación de la Calidad de Vida es, pues, una valoración personal.

Por instrumento entenderemos el método por el cual se consiguen datos, cualitativos o cuantitativos, sobre los pacientes: cuestionarios, entrevistas, procedimientos observacionales, y cualquier otro material rellenado por el paciente o el investigador.

Un test es un procedimiento sistemático utilizado para observar algún aspecto del comportamiento humano y describirlo mediante una escala numérica o un sistema de categorías.

En ocasiones el término factor es usado como sinónimo de constructo, pero esta expresión es también utilizada para denotar un nivel más bajo (componentes). En este sentido, un constructo, por ejemplo, Calidad de Vida, puede ser descompuesto en varios componentes o factores, cada uno de ellos representando un dominio o dimensión. Cada dimensión se refiere al área que estamos tratando de medir. Así, actividad física, entorno social y estado emocional son variables latentes, las cuales son todas ellas aspectos o factores en Calidad de Vida.

¹ En inglés para medir se usa el término "measurement", pero para medir Calidad de Vida utilizan el término "assessment".

Debido a que algunos de los componentes de la Calidad de Vida no pueden ser observados directamente, éstos se evalúan a través de tests que constan de un conjunto de *ítems* que son preguntas, cuyas respuestas (assessments) se realizan en una *escala* nominal u ordinal, generalmente. Cada pregunta representa una variable que aporta un peso específico a una cuantificación global, para un factor o dominio.

Para referirnos a cualquier método de medida o valoración emplearemos el término test o cualquier sinónimo, escala, cuestionario, etc.

En teoría, se asume que hay un valor verdadero para la Calidad de Vida y que puede medirse indirectamente por medio de tests. Cada variable mide un concepto, y combinadas conforman una escala estadísticamente cuantificable.

Para subrayar estas ideas descritas, consideremos un ejemplo en el que la variable latente es la Calidad de Vida en pacientes con Osteoporosis. El QUALEFFO (BADIA y HERDMAN, 1999), es el cuestionario más utilizado en medicina contemporánea para medir la Calidad de Vida en pacientes con osteoporosis, consta de 35 ítems las cuales se refieren a siete dimensiones de salud: Dolor (5 ítems), Actividades Cotidianas (3 ítems), Tareas Domésticas (5 ítems), Movilidad (6 ítems), Actividades Sociales y de Tiempo Libre (4 ítems), Percepción de la Salud General (2 ítems) y Estado de Ánimo (10 ítems). (Figura 1.3).

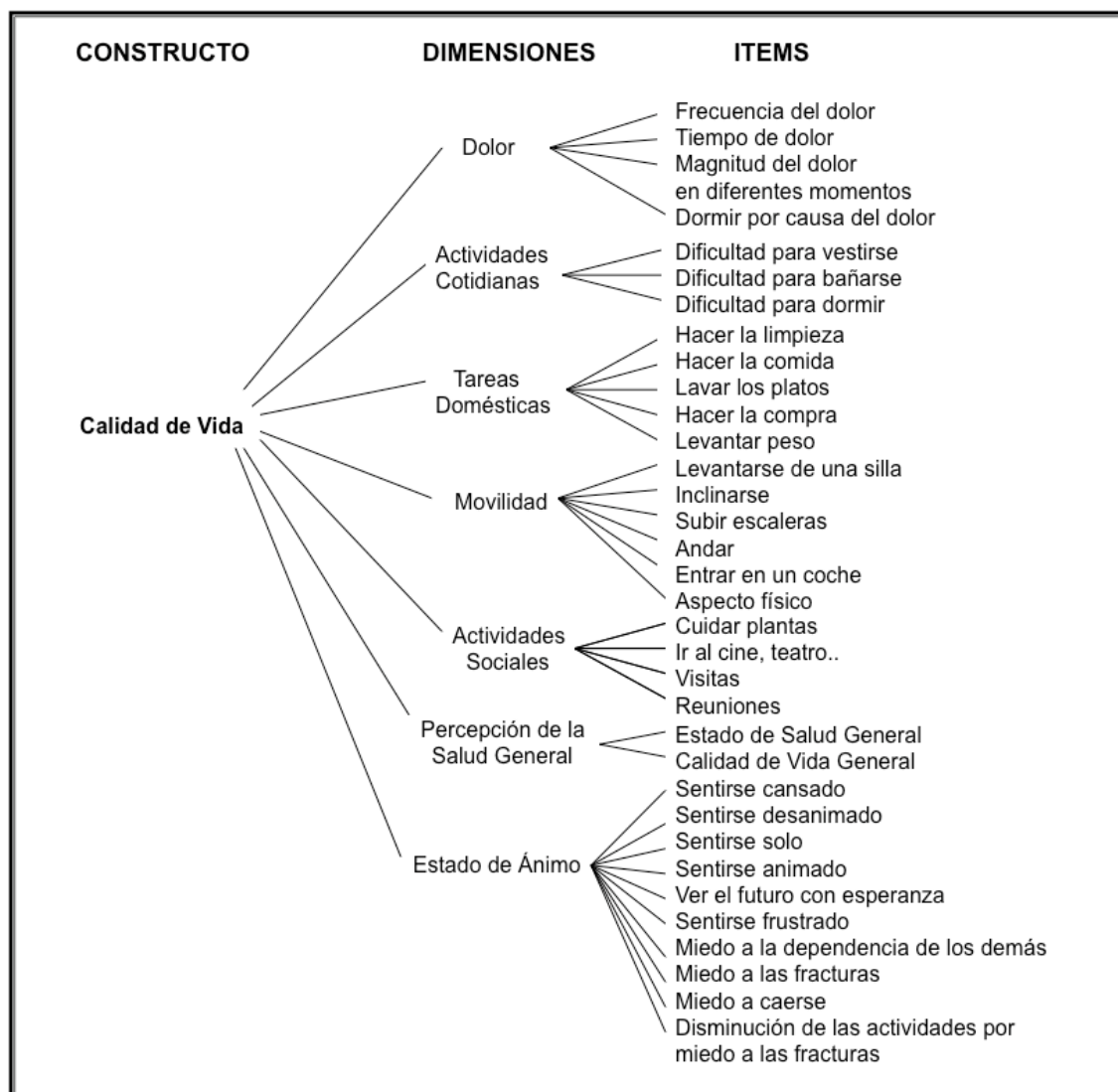


Figura 1.3.- Estructura del Cuestionario QUALEFFO

Las medidas de CVRS vienen en una variedad de formas que reflejan su uso. La principal distinción es entre un *índice* y un *perfil*. Los *perfiles* consisten en múltiples escalas que reflejan las múltiples dimensiones de la Calidad de Vida como la física, emocional, funcional, y el bienestar social de los pacientes. En la mayoría de los instrumentos, cada escala se construye a partir de las respuestas a varias preguntas (ítems). Un perfil de los diversos dominios refleja el carácter multidimensional de la Calidad de Vida (AARONSON y col., 1992).

Un *índice* es una puntuación única global que ordena a los individuos en función de la forma en que perciben la salud (NUNNALLY y BERNSTEIN, 1994).

Existen dos métodos para la creación de índices; en el primero se utiliza una única pregunta para evaluar la Calidad de Vida del paciente. El segundo método consiste en proporcionar, por parte de los investigadores, métodos para combinar las respuestas a múltiples preguntas para obtener un único índice de Calidad de Vida.

La ventaja de obtener un índice único es que proporciona un enfoque sencillo a la toma de decisiones, como los ensayos clínicos, donde la Calidad de Vida es el resultado primario. Es imposible construir un índice que agrega las múltiples dimensiones de la CVRS que será adecuado en todos los contextos.

1.2.2.- CÓMO MEDIR CVRS

Las variables latentes y los constructos son conceptos abstractos (*“useful fictions”* en palabras de FAYERS y MACHIN (2000a). No pueden ser medidos directamente y son evaluados por medio de tests de múltiples ítems y cuestionarios. Por el contrario, las denominadas variables manifiestas son las respuestas observadas y realizadas por el paciente a los ítems del cuestionario.

Las dimensiones latentes que son importantes para la medición de la CVRS son *“el funcionamiento social, físico, y cognitivo; la movilidad y el cuidado personal; y el bienestar emocional”* (HERDMAN y BARÓ, 2000).

Formas de respuesta

La mayoría de los instrumentos de CVRS recogen las respuestas en escalas de tipo Likert. Estas cuestiones son planteadas al paciente, y las respuestas son transformadas en puntuaciones numéricas que se combinan para proporcionar puntuaciones del dominio.

En el cuestionario para medir Calidad de Vida en pacientes con Osteoporosis (QUALEFFO), todas las preguntas se expresan en escala politómica (Ver Figura 1.4), unas tienen cinco categorías y otros ítems seis categorías de respuesta. (Ver Anexo 4).

Durante los últimos siete días, ¿cómo ha sido el dolor de espalda en el peor de los casos?

1 ☐ No he tenido dolor de espalda 2 ☐ Suave 3 ☐ Moderado 4 ☐ Fuerte 5 ☐ Insoportable

Figura 1.4.- Ítem del Cuestionario QUALEFFO

También se pueden presentar formatos de respuesta dicotómica (presencia o ausencia) y Escala Visual Analógica (EVA). La EVA (Figura 1.5) consiste en una línea, de unos 10 cm de longitud, cuya posición representa, de abajo a arriba el valor de la Calidad de Vida del paciente. La motivación original de la EVA fue el hecho de pensar que la medida continua podría discriminar de manera más eficaz que una escala Likert, lo que no ha sido generalmente cierto en la mayoría de los estudios de validación en los que se han utilizado ambos formatos.

La EVA requiere un nivel de coordinación “mano-ojo” que puede ser problemático para personas con problemas neurológicos, para individuos que experimentan entumecimiento y hormigueo, para pacientes que sufren los efectos secundarios de la quimioterapia, o para las personas mayores. A pesar de estas limitaciones es la medida que se utiliza usualmente para medir CV general.

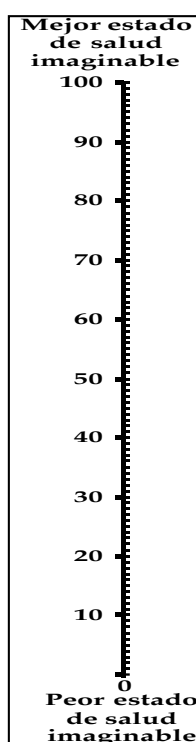


Figura 1.5.- Escala Visual Analógica

Como veremos más adelante, otro grupo de instrumentos para medir la Calidad de Vida son las medidas basadas en la utilidad. Estas medidas vienen siempre expresadas en una escala de 0 a 1. La puntuación depende del método utilizado para obtener las preferencias.

1.2.3.- PERIODO DE EVALUACIÓN DE LA CVRS

Las escalas de medición de Calidad de Vida basan su evaluación durante un período de tiempo determinado como, “durante los últimos siete días” o “durante las últimas cuatro semanas”. Este período de tiempo debe ser lo suficientemente corto para detectar diferencias entre los tratamientos y el tiempo suficiente para minimizar las fluctuaciones de corto plazo (ruido) que no representan el verdadero cambio.

Escalas específicas para ciertas enfermedades, o para ciertos tratamientos, donde puede haber cambios rápidos en el periodo de evaluación, necesitarán un tiempo menor que en las evaluaciones de Calidad de Vida general. Por ejemplo, los ítems de la dimensión *Dolor* del cuestionario específico QUALEFFO, para medir Calidad de Vida en pacientes con osteoporosis, hacen una evaluación del dolor y de la intensidad del dolor de espalda percibido por el paciente durante los *siete días anteriores* al de la realización de la encuesta.

Por citar otro ejemplo, en el SF-36 (WARE y col., 1992) que es el cuestionario genérico sobre Calidad de Vida Relacionada con la Salud más utilizado en la medicina contemporánea, algunos de sus ítems hacen referencia a la evaluación del estado de salud del paciente *en el momento actual*, otros sobre *las últimas cuatro semanas* e incluso otros ítems piden hacer una comparación del estado actual del paciente en el momento en el que se está realizando la encuesta con el estado de salud de *un año anterior* a ese momento.

1.2.4.- INSTRUMENTOS PARA MEDIR CVRS

El interés por medir la Calidad de Vida Relacionada con la Salud ha dado lugar al desarrollo de un número considerable de instrumentos de medida, cuya finalidad es aportar valoraciones más precisas de la salud de los individuos o de las poblaciones y de los beneficios y perjuicios que pueden resultar de la atención sanitaria.

Hay dos grandes tipos de instrumentos para la medición de la Calidad de Vida Relacionada con la Salud utilizados en la práctica asistencial y en los ensayos clínicos: los cuestionarios y las medidas basadas en la utilidad. A su vez, los cuestionarios pueden ser genéricos o específicos. El desarrollo de estos dos tipos de instrumentos es el resultado de las diferencias entre las perspectivas de dos disciplinas diferentes: la Psicometría y la Econometría.

1.2.4.1.- CUESTIONARIOS

Los cuestionarios exploran separadamente las distintas dimensiones o aspectos del daño que ocasionan las enfermedades y su propósito es fundamentalmente analítico, ya que pretenden conseguir información diferenciada de los distintos atributos de la Calidad de Vida.

En la evaluación de las medidas del estado de salud, múltiples aspectos del “bienestar del paciente” son percibidos y de ello se deriva una puntuación de las respuestas a una serie de preguntas. Esta puntuación refleja la CVRS del paciente en comparación con otros pacientes y en la CVRS del mismo paciente en otros momentos. Los cuestionarios están diseñados principalmente para comparar grupos de pacientes que reciben tratamientos diferentes o para identificar cambios en el tiempo dentro de los mismos grupos de pacientes.

Entre estas medidas del estado de salud, existe una gran variedad de preguntas; unas centrándose más en la percepción del impacto de la enfermedad y la terapia (¿En que medida le afecta estar más encorbado?), otras medidas centradas en la frecuencia y la gravedad de los síntomas (¿Con qué frecuencia le duele la espalda?), y otras más

generales de la evaluación de la Calidad de Vida (Teniendo en cuenta su edad, ¿Cómo valoraría su Calidad de Vida?).

Debido a que muchos de los dominios y componentes de la Calidad de Vida no pueden ser observados directamente, éstos son evaluados de acuerdo a los principios clásicos de la teoría ítem-medida (LORD, 1980). Así, las medidas de perfil son instrumentos en forma de cuestionarios que contienen múltiples ítems o preguntas que pueden ser formuladas o contestadas por uno mismo (cuestionario autoadministrado) o bien consignadas por un entrevistador profesional.

Los perfiles de salud pueden aplicarse como herramienta de encuesta para poblaciones seleccionadas, como complemento a la historia clínica, en la evaluación de tratamientos médicos, para ensayos clínicos con grupos seleccionados de pacientes y como instrumento de ayuda en la toma de decisiones en política sanitaria (CASAS ANGUITA y col., 2001).

Para algunos cuestionarios es posible obtener a partir de las respuestas dadas un único índice o resumen, agrupando las puntuaciones de las diferentes dimensiones en una puntuación única o global. Esta puntuación única puede servir para medir o comparar de forma sencilla distintos cursos de acción o estados de salud.

Cuando se utiliza un cuestionario para evaluar la Calidad de Vida Relacionada con la Salud, interesa establecer si el interés radica en evaluar distintas actuaciones para tratar una enfermedad específica (esencialmente para comparar distintos cursos de acción ante un problema de salud), o bien si lo que se quiere es analizar de forma más amplia el daño o carga de enfermedad que ocasionan las enfermedades (lo que se orienta tácitamente a comparar estados de salud entre sí para el establecimiento de prioridades sociales de intervención). Así, los instrumentos de medición pueden ser genéricos o específicos.

Instrumentos genéricos

Los instrumentos genéricos son aquéllos que son aplicables a cualquier tipo de población independientemente de su condición demográfica, clínica o terapéutica. Proporcionan el perfil de salud de un individuo o de una población según las diferentes dimensiones de la Calidad de Vida.

El amplio contenido de los ítems de un instrumento genérico es una ventaja cuando se comparan diferentes grupos de sujetos o se hace un seguimiento durante largos períodos de tiempo después de que el tratamiento ha terminado.

Otra ventaja de un instrumento genérico es que permite establecer comparaciones entre distintas enfermedades y, proporciona por tanto, información útil en la planificación y gestión de recursos, así como comparaciones sobre la Calidad de Vida entre diversas poblaciones de pacientes.

Presentan el inconveniente de que, al ser de uso universal, los ítems son demasiado inespecíficos y globales y pueden ser considerados por los pacientes como no adecuados a su situación particular; asimismo, pueden no ser lo suficientemente sensibles como para detectar cambios en el tiempo que sean significativos para los pacientes de determinadas enfermedades (GUYATT y col., 1997).

Entre los más utilizados y adaptados para su uso en España, encontramos:

A) El Perfil de las Consecuencias de la Enfermedad (Sickness Impact Profile, SIP). Fue diseñado por MARILYN BERGNER y col. entre 1972 y 1976 para analizar los cambios en el comportamiento de una persona y el grado de disfunción debido a una enfermedad. Es una de las medidas de Calidad de Vida más conocidas y más usadas y es una medida general aplicable a cualquier grupo de enfermedad o discapacidad (BERGNER y col., 1981).

La base conceptual de este cuestionario estriba en la percepción que el propio paciente tiene de su enfermedad a través de su efecto sobre las actividades de la vida diaria, los sentimientos y las actitudes del sujeto.

Este concepto se diferenció en la propia denominación de la escala, donde se incluyó la palabra “sickness” (sentirse enfermo) en lugar de “disease” (enfermedad) (BERGNER y col., 1981).

El SIP contiene 136 ítems agrupados en 12 categorías: deambulación (12 ítems), cuidado personal/movimiento corporal (23 ítems), movilidad (10 ítems), tareas domésticas (10 ítems), ocio/pasatiempos (8 ítems), interacción social (20 ítems), comportamiento emocional (9 ítems), conducta con atención al medio (10 ítems), sueño y descanso (7 ítems), comida (9 ítems), comunicación (9 ítems) y trabajo (9 ítems). Siete de estas categorías pueden agruparse en dos dimensiones, física y psicosocial, mientras que las cinco restantes son independientes (sueño y cansancio, comida, trabajo, ocio, pasatiempos, tareas domésticas). La dimensión física está integrada por tres categorías: desplazamiento, movilidad y cuidado personal/movimiento corporal. La dimensión psicosocial está integrada por cuatro categorías: interacción social, conducta con atención al medio, comportamiento emocional y comunicación (BERGNER y col., 1981).

El cuestionario puede ser autoadministrado o aplicado por un entrevistador y el paciente debe señalar aquellos ítems que describen su estado de salud actual y que están relacionados con su enfermedad principal. Cuando es administrado se invierten de 20 a 30 minutos. Se requieren otros 10 minutos más por cuestionario para interpretar las puntuaciones.

Cada ítem del SIP ha sido evaluado por un grupo independiente de jueces sobre una escala de disfunción de 15 puntos. Las valoraciones de los jueces determinan la ponderación de cada ítem en la puntuación total. El paciente no considera las ponderaciones de los jueces cuando decide marcar, o no, un ítem. La puntuación porcentual global se obtiene seleccionando los ítems marcados por el paciente, sumando sus valores y dividiendo el resultado obtenido por la suma de los valores de todos los ítems del SIP. Después, esta proporción se multiplica por 100 para calcular una puntuación porcentual total de afectación. Similarmente, pueden obtenerse puntuaciones porcentuales de cada categoría que pueden ser representadas en una gráfica.

Existen dos versiones adaptadas al español realizadas por VÁZQUEZ-BARQUERO y col. (1991) y BADÍA y ALONSO (1995) respectivamente.

B) El estudio de los resultados médicos a través del formulario breve de 36 ítems (SF-36): Fue diseñado por WARE y colaboradores como una medida genérica del estado de salud para la utilización en muestras de la población y en estudios de política sanitaria (WARE y col., 1992; McHORNEY y col., 1993; McHORNEY y col., 1994). Se desarrolló para ser empleado en el Estudio de Resultados Médicos (Medical Outcomes Study, MOS), a partir de una extensa batería de cuestionarios que incluían 149 conceptos de la salud aplicados en casi 22.000 pacientes de EEUU.

Es la medida del estado de salud usada más habitualmente en el mundo actual. El formato final es un instrumento genérico compuesto por 36 ítems que pretenden medir ocho conceptos genéricos sobre la salud (esto es, conceptos que no son específicos de una enfermedad, grupo de tratamiento o edad), detectando tanto estados positivos como negativos de la salud física y el estado emocional.

Los ocho conceptos de salud (dimensiones) determinados en este cuestionario se resumen en la siguiente tabla (Tabla 1.1).

Dimensión	Significado
Función física (10 ítems)	Grado en el que la falta de salud limita las actividades físicas de la vida diaria, como el cuidado personal, caminar, subir escaleras, coger o transportar cargas, y realizar esfuerzos moderados e intensos.
Rol físico (4 ítems)	Grado en el que la falta de salud interfiere en el trabajo y otras actividades diarias, produciendo como consecuencia un rendimiento menor del deseado, o limitando el tipo de actividades que se puede realizar o la dificultad de las mismas.
Dolor corporal (2 ítems)	Medida de la intensidad del dolor padecido y su efecto en el trabajo habitual y en las actividades del hogar.
Salud general (5 ítems)	Valoración personal del estado de salud, que incluye la situación actual y las perspectivas futuras y la resistencia a enfermar.
Vitalidad (5 ítems)	Sentimiento de energía y vitalidad, frente al de cansancio y desánimo.
Función social (2 ítems)	Grado en el que los problemas físicos o emocionales derivados de la falta de salud interfieren en la vida social habitual.
Rol emocional (3 ítems)	Grado en el que los problemas emocionales afectan al trabajo y otras actividades diarias, considerando la reducción del tiempo dedicado, disminución del rendimiento y del esmero en el trabajo.
Salud mental (5 ítems)	Valoración de la salud mental general, considerando la depresión, ansiedad, autocontrol, y bienestar general.

Tabla 1.1.- Ítems del cuestionario SF-36

Las mediciones sobre la Calidad de Vida con la escala SF-36 no parecen verse modificadas o influenciadas por la forma de administración del cuestionario (entrevista personal, telefónica o autoadministrada).

ALONSO y col. (1995) han adaptado y traducido la versión española del SF-36 bajo el nombre de cuestionario de salud SF-36 y han establecido valores poblacionales de referencia por sexo e intervalos de edad.

C) El Perfil de Salud de Nottingham (Nottingham Health Profile, NHP): Es un cuestionario de medida de salud percibida, desarrollado originalmente en Inglaterra por HUNT y col. (1981). La versión original consta de 45 ítems divididos en dos apartados. El primero, con 38 ítems, valora seis dimensiones o situaciones causadas por problemas de salud: energía (3 ítems), dolor (8 ítems), movilidad física (8 ítems), reacciones emocionales (9 ítems), sueño (5 ítems), aislamiento social (5 ítems). El segundo

contiene 6 cuestiones acerca de la presencia o no de limitación por causa de la salud en las áreas de la vida afectadas por la salud: empleo, actividades del hogar, vida social, vida familiar, vida sexual, *hobbies* e intereses y vacaciones. Puede ser autoadministrado por el propio paciente en 5-10 minutos. La versión original ha sido traducida y validada en nuestro país con el nombre de Perfil de Salud de Nottingham.

El NHP tiene muchas características atrayentes. La escala está basada en el usuario y procede de definiciones de salud dadas por individuos de la colectividad. Además, el NHP fue diseñado con un lenguaje fácilmente entendido por el común de las personas y que exige el menor número posible de requisitos de comprensión. La principal dificultad con el NHP es que no proporciona ponderaciones de la importancia relativa a lo largo de las dimensiones. Como consecuencia, es difícil comparar las dimensiones unas con otras.

D) El EuroQol, también llamado EQ-5D: Fue desarrollado en el marco de un proyecto europeo común que comenzó en 1987, iniciado entre otros, por el profesor ALAN WILLIAMS en la Universidad de York en Inglaterra.

Es un instrumento estandarizado para el uso de la medida de la Calidad de Vida Relacionada con la Salud. Se ha diseñado especialmente para completar las medidas con el SF-36, el NHP, el SIP o los instrumentos específicos.

Este cuestionario (EQ-5D) consta de cuatro partes:

1ª.- Descripción del propio estado de salud en cinco dimensiones (movilidad, cuidado personal, actividades cotidianas, dolor/malestar y ansiedad/depresión), codificándose cada una de ellas con un 1 cuando no se perciben problemas, con un 2 cuando se estiman algunos/moderados y con un 3 cuando los problemas son muchos. El individuo debe marcar el nivel de problemas, dentro de cada dimensión, que mejor describe su estado de salud “en el día de hoy”. De este modo, se define el estado de salud del individuo mediante un número de cinco dígitos. El mejor estado de salud se describe 11111 y el peor 33333. Son posibles $3^5=243$ estados de salud.

2ª.- Esta parte consiste en una Escala Visual Analógica (EVA), vertical y milimetrada de 20 cm de longitud, semejante a un termómetro, en la que sus extremos aparecen etiquetados como “el peor estado de salud imaginable” y “el mejor estado de salud imaginable”, con puntuaciones de 0 y 100, respectivamente. El individuo debe señalar en la escala el punto que, en su opinión, indica su estado de salud en el día de hoy. El valor asignado en la EVA puede utilizarse como un indicador cuantitativo del estado de salud de un individuo o de un grupo de individuos.

3ª.- Esta parte pretende alcanzar la valoración individual de los estados de salud definidos por el sistema descriptivo del EQ-5D; incluye además los de “inconsciente” y “muerte”. La valoración individual de los estados de salud se presenta en dos páginas consecutivas. En cada página aparece una EVA en la parte central con ocho estados de salud descritos en recuadros a ambos lados de ella. Se les pide a los encuestados que imaginen que se hallan en cada estado de salud y que unan mediante líneas cada uno de los recuadros con un valor de la EVA, indicando así su preferencia por cada uno de los estados de salud. Como test de consistencia se repiten el mejor y el peor estado posible. Una vez asignados los valores a los 16 estados de salud presentados, se pide a los encuestados que asignen un valor a la “muerte” sobre la misma EVA en ambas páginas. A partir de los valores obtenidos para los estados de salud se puede obtener, utilizando modelos de regresión, una valoración social que se representa en una tarifa de valores para los 243 estados de salud, además de los estados “inconsciente” y “muerte”. Para calcular los valores de la tarifa se pueden utilizar datos no ajustados (asignación directa de los valores obtenidos en la EVA) o ajustados (los estados “11111” y “muerte” toman valores 1 y 0, respectivamente; con esta referencia, son ajustados los restantes datos).

4ª.- Finalmente, se recoge información del encuestado: edad, sexo, experiencia en enfermedades (propia, en familiares y en otras personas), dificultad para cumplimentar el cuestionario y experiencia en completar cuestionarios similares.

Este instrumento, de utilización creciente, fue revisado en 1996 y adaptado al español por BADÍA y col. (1996 y 1999).

Instrumentos específicos

Los instrumentos específicos son aquéllos que han sido específicamente diseñados para ser aplicados a una población concreta, teniendo en cuenta sus características demográficas, clínicas o terapéuticas. Estos instrumentos no permiten, por lo tanto, la comparación del nivel de Calidad de Vida entre distintos grupos. Al ser tan específicos presentan la ventaja de que sus ítems son relevantes para la condición para la que han sido desarrollados y son vistos por los pacientes como muy pertinentes, ya que abordan la problemática concreta de su situación.

Los instrumentos específicos suelen reducir la sobrecarga de administración del cuestionario para el paciente y suelen aumentar la aceptabilidad, al incluir sólo dimensiones relevantes, lo que puede hacer aumentar, por otro lado, la sensibilidad al cambio.

Los instrumentos específicos pueden ser específicos de enfermedad (por ejemplo, asma), de una población de pacientes (tales como ancianos frágiles), de una función (tal como sueño) o de un problema (por ejemplo, dolor).

Entre los instrumentos específicos para los enfermos de osteoporosis se encuentran el QUALEFFO y el OQLQ. Estos cuestionarios han sido adaptados al español por BADÍA y HERDMAN en 1999.

El OQLQ (Osteoporosis Quality of Life Study Group) es un cuestionario que se administra mediante entrevista y que consiste en 30 ítems que cubren cinco dimensiones de la Calidad de Vida Relacionada con la Salud: 9 ítems sobre síntomas, 5 ítems sobre funcionamiento emocional, 9 ítems sobre actividades cotidianas, 4 ítems sobre funcionamiento emocional y 3 ítems sobre actividades sociales y de ocio. La respuesta a cada ítem está en escala ordinal, yendo desde 1 (muchos problemas) a 7 (ningún problema). La puntuación para cada dimensión se obtiene sumando los valores de las respuestas a cada ítem dentro de una dimensión y dividiendo por el número de ítems en la dimensión, proporcionando así una puntuación media para cada dimensión de manera que una puntuación más alta corresponde a una mejor Calidad de Vida Relacionada con la Salud.

La versión adaptada del QUALEFFO (Questionnaire of the European Foundation for Osteoporosis) difiere ligeramente del cuestionario original incluyó finalmente 35 ítems , frente a los 41 ítems del original, que preguntaban sobre 7 dimensiones de la Calidad de Vida Relacionada con la Salud: 5 ítems sobre dolor, 3 ítems sobre actividades cotidianas, 5 ítems tareas domésticas, 6 ítems sobre movilidad, 4 ítems sobre actividades sociales, 2 ítems sobre la Calidad de Vida Relacionada con la Salud global y 10 ítems sobre el funcionamiento mental. La respuesta a cada ítem está también en escala ordinal, yendo desde 1 (ningún problema) a 5 (muchos problemas). El cuestionario fue diseñado para ser autoadministrado (LIPS y col., 1997).

Interpretación de los cuestionarios

El concepto “interpretar” es muy importante a la hora de utilizar los cuestionarios ya que éstos se diseñan para unos objetivos concretos, que hay que conocer si se quiere obtener conclusiones correctas (MOLINERO, 1998).

Existen dos métodos para interpretar un cuestionario, la interpretación de tipo descriptivo del estado de salud de un paciente concreto o de un conjunto de individuos con unas características determinadas. O puede ser una media comparativa entre dos grupos de pacientes, por ejemplo con diferentes tratamientos.

Siempre se ha considerado que la evolución de una enfermedad va íntimamente ligada a las "ganas de vivir" que le quedan al paciente, lo cual no se sabe si es la causa o el efecto, pero en cualquier caso ese concepto vitalista o humanístico subyace en el objetivo de las escalas de medida de la Calidad de Vida.

En un cuestionario se pueden interpretar las puntuaciones obtenidas, con el fin de poder obtener conclusiones que puedan ayudar a la toma de decisiones clínicas, económicas o incluso diagnósticas.

No conviene olvidar que estamos manejando valores subjetivos, que a diferencia de una analítica (registro de presión arterial, una radiografía, etc.) son más difíciles de medir, y que se pueden ver alterados por factores como el número de preguntas, o la comprensibilidad de éstas, que influyen en el tiempo de cumplimentación, van a alterar

el grado de cooperación del paciente y por consecuencia el rigor de la información recogida.

Entre las diferentes técnicas estadísticas a la hora de interpretar los resultados de una encuesta, tenemos por ejemplo, la comparación de los cambios en los pacientes en diferentes estadios de la enfermedad, para el caso de cuestionarios específicos, o para los cuestionarios que miden Calidad de Vida global, comparar las puntuaciones medias de los pacientes con diferentes patologías.

En el caso de los cuestionarios genéricos, como es el SF-36, el método más ampliamente utilizado es realizar un estudio de una muestra representativa de la población general, con el fin de obtener valores que puedan ser considerados normales en la población general, para poder contrastar con esos valores de referencia los resultados de cada estudio.

Evaluación de los cuestionarios

BOMBARDIER y TUGWELL (1987) hicieron una revisión sobre los líneas de estudio que sirven de referencia para diseñar un instrumento para medir Calidad de Vida.

Objetivo: Se debe evaluar ¿qué enfermedad se va a estudiar?, ¿en qué población se va a aplicar?.

Se recomienda usar instrumentos ya existentes, ya que hay que tener en cuenta que la elaboración y validación de un instrumento no garantiza que sea útil.

Función: El instrumento debe de definirse en función de su capacidad de discriminación (para establecer diferencias entre pacientes con la misma enfermedad o entre padecimientos), en función de su capacidad descriptiva (debe de ser capaz de describir la evolución de una enfermedad y detectar cambios de Calidad de Vida a través del tiempo) y en función de su capacidad de predecir (predicción del tipo de discapacidad futura y un desenlace poco afortunado).

Selección de preguntas: Se deben definir preguntas relevantes para medir la Calidad de Vida en pacientes con una determinada enfermedad. En un cuestionario muchas preguntas pueden estar correlacionadas y tener equivalencias. Se debe de eliminar la redundancia y medir una dimensión con el mínimo número de variables. Las técnicas estadísticas más empleadas para reducir y caracterizar las dimensiones son el análisis factorial y las componentes principales.

Contenido: Deben incluirse dimensiones importantes, como por ejemplo, tareas domésticas, movilidad, cuidado personal, etc., cuando se pretende medir Calidad de Vida en pacientes con osteoporosis.

Formas de respuesta: Se pueden contestar de diferentes formas de manera que al sumar las calificaciones se pueda obtener una evaluación representativa del área que se quiere medir.

Sentido biológico: En la medida en que los instrumentos cumplan con las predicciones clínicas lograrán mayor credibilidad. Estas mediciones representan un complemento para el cuidado integral del paciente.

Factibilidad: Los instrumentos deben de ser adecuados para las diferentes culturas en las cuales se aplican. El tiempo que se consume en su aplicación puede ser limitada en la práctica clínica (PATRICK y DEYO, 1989 y CARDIEL, 1994).

En la siguiente tabla se presenta una síntesis de las características de los instrumentos descritos anteriormente, con información en relación con su objetivo, dimensiones y función.

INSTRUMENTO	OBJETIVO	DIMENSIONES	FUNCIÓN
SIP	Validación	Funcionalidad físico, emocional,	Discriminación
	Confiabilidad	social	Evaluación
	Traducción al español		
SF-36	Validación	Funcionalidad físico, social,	Discriminación
	Confiabilidad	emocional, dolor	Evaluación
NHP	Validación	Funcionalidad físico, emocional,	Discriminación
	Traducción al español	social, nivel actividad, percepción deterioro	
EUROQOL-5D	Validación	Funcionalidad físico, autocuidado,	Discriminación
	Confiabilidad	ansiedad, limitación	

Tabla 1.2.- Evaluación de instrumentos para valorar Calidad de Vida, tomada de VELARDE-JURADO y AVILA-FIGUEROA (2002)

Traducción y adaptación transcultural de los cuestionarios

Si un instrumento de medida ha sido desarrollado originalmente en un contexto cultural, o una población diana diferente a la que se va a aplicar, es necesario adaptar el cuestionario mediante un proceso de adaptación cultural que está establecido internacionalmente, el cual ha de preservar el contenido semántico en el uso habitual del lenguaje de la nueva traducción, intentando lograr la equivalencia de significados del cuestionario en ambas culturas.

El cuestionario QUALEFFO se adaptó culturalmente al español mediante el proceso de traducción-retrotraducción de la versión original hasta obtener una versión final consensuada y realizar una prueba piloto para evaluar su comprensión y aceptabilidad en una muestra de pacientes. También ha sido adaptado recientemente en otros países, por ejemplo, a la población Mexicana (RAMIREZ y col., 2007) y a la población Turca (KOCYIGIT y col., 2003).

Cuando se mide CVRS en distintas poblaciones, es preciso prestar atención a los métodos de traducción y a las características culturales. Hay numerosos trabajos en los que los investigadores han encontrado problemas en la validación y traducción de cuestionarios en diferentes idiomas y culturas (ZASLAVSKY, 2005). La Teoría de Respuesta al Ítem y los modelos de RASCH facilitan el examen del Funcionamiento Diferencial de los Ítems a través de diferentes culturas o lenguas.

La mayoría de los instrumentos disponibles para medir Calidad de Vida han sido desarrollados en inglés, por lo que su aplicación en países de habla hispana la traducción de un instrumento de medición se realiza en cuatro fases:

En la primera fase, se traduce el documento original del inglés al español; esta traducción debe ser realizada por investigadores involucrados en el estudio.

En la fase dos, la versión en español obtiene validez de consenso, validez aparente y validez de contenido, después de que cada uno de los ítems es sometido a revisión por profesionales con experiencia en la atención a pacientes con la enfermedad en estudio.

En la fase tres, la versión en español se entrega a un traductor oficial, no relacionado con el equipo de trabajo, para obtener una nueva versión de español a inglés.

En la fase cuatro, el documento original y la traducción hecha por un perito oficial (ambos en inglés), son sometidos a comparación para verificar que no existen discrepancias graves en la versión original y la traducción (GANDEK y WARE, 1998).

1.2.4.2.- MEDIDAS DE UTILIDAD

Los clínicos hacen interpretaciones generales del perfil aplicando un sistema de pesos. El juicio acerca de la importancia relativa de varias dimensiones es habitualmente realizado implícita y arbitrariamente y de un modo idiosincrásico (KAPLAN, 1997).

Por otra parte, en ocasiones es difícil recoger el cuadro clínico global usando medidas de perfil debido a que la mayoría de los tratamientos tiene tanto efectos secundarios como beneficios.

Además de los problemas al interpretar perfiles para la toma de decisiones clínicas, es difícil usar perfiles para el análisis coste-utilidad. La comparación de diferentes opciones para el uso de recursos comunes requiere la cuantificación de resultados de salud usando una unidad de medida común: los años de vida ajustados por la calidad (AVACs). Los años de vida ajustados por la calidad integran la mortalidad y la

morbilidad para expresar el estado de salud en términos equivalentes a años de vida con buena salud. Por tanto, tienen en cuenta las consecuencias de las enfermedades sobre la Calidad de Vida, considerando tanto los beneficios como los efectos secundarios de los tratamientos.

El análisis coste-utilidad es un análisis en el que se expresan los efectos de una intervención en términos de preferencias personales (utilidades) y se describen los costes para alguna ganancia adicional en utilidades (por ejemplo, coste por un año de vida ajustado por la calidad o AVAC adicional).

Esta clase de evaluación se emplea fundamentalmente en estudios en donde la variación en la Calidad de Vida es uno de los resultados más importantes del estudio, afectándose o no la cantidad de la misma. Un ejemplo de ello es el estudio farmacoeconómico de los parches transdérmicos de estradiol en la mujer menopáusica. Esta alternativa puede disminuir la cantidad de vida debido al aumento de la mortalidad por mayor prevalencia del cáncer de endometrio, pero puede aumentar la Calidad de Vida de la mujer en función de una disminución de la sintomatología propia de la menopausia. El análisis coste-utilidad decidiría, al pasar ambos tiempos a unidades homogéneas (años de vida ajustados por la calidad), cuál de las dos posibilidades (tratar o no con parches) ofrece una solución más satisfactoria para la paciente.

Así, este tipo de análisis debería tenerse en cuenta para decidir entre varias opciones terapéuticas de una misma enfermedad o bien podría servir para favorecer la financiación pública de un tratamiento frente a otro u otros que demuestren una menor relación coste-utilidad.

El ajuste sobre la Calidad de Vida se realiza en función de las utilidades sobre los estados de salud. De manera general, las utilidades son valores cardinales que reflejan la deseabilidad o la fuerza de las preferencias individuales para resultados específicos bajo condiciones de incertidumbre. Las utilidades de la salud son las preferencias para estados de salud específicos o para tratamientos.

Las cuestiones que deben tratarse en cualquier estudio de utilidad incluyen asegurar que los atributos y niveles relevantes de salud son incorporados en el instrumento de medida, determinar el método de ordenación o escalamiento de preferencia apropiado para obtener las utilidades, decidir los sujetos adecuados cuyas preferencias debieran ser evaluadas y minimizar los efectos de contexto sobre los resultados.

a) *¿Cuáles son las dimensiones relevantes de la salud?*

El componente crucial más simple de cualquier estudio que mida la Calidad de Vida Relacionada con la Salud es asegurar que las dimensiones y niveles relevantes de la salud son incorporados en el instrumento de medida. En este campo de la medición de la utilidad, las dimensiones de la salud se denominan atributos. Algunos ejemplos de atributos de la salud son la función física, social y cognoscitiva, el bienestar psicológico y diversos síntomas (dolor, disnea, astenia, etc.). En general, el número máximo de atributos que pueden incluirse en un estado de salud es nueve (preferiblemente menos), basándose en la investigación que indica que los seres humanos pueden procesar simultáneamente sólo nueve piezas de información (MILLER, 1956).

b) *¿Qué método de ordenación de preferencias debe usarse?*

No hay acuerdo sobre cómo deben medirse las utilidades. En la mayoría de los enfoques, los encuestados asignan pesos a los diferentes estados de salud en una escala que va desde 0 (para la muerte) a 1 (para la salud plena). Las técnicas más importantes para valorar utilidades son la jugada estándar, el intercambio temporal, la escala de puntuación, la estimación de la magnitud, el método de equivalencia y la técnica de la voluntad de pago. De ellas nos ocuparemos a continuación.

La **jugada estándar** (*Standard gamble*) es el método clásico, originado en el campo de la teoría de la decisión, de medir las preferencias. Propuesto por von NEUMANN y MORGENTERN (1953), está basado en los axiomas de la teoría de la utilidad y comprende un marco conceptual para examinar la toma de decisiones bajo incertidumbre. La esencia de la técnica es una elección planteada al interrogado entre un resultado cierto y un juego. La Figura 1.6 ilustra la jugada estándar para un estado de salud crónico preferido a la muerte. La alternativa 1 es un juego con dos posibles resultados: o el paciente recupera su salud normal y vive t años más (probabilidad p) o

el paciente muere inmediatamente (probabilidad $1-p$). La alternativa 2 consiste en el resultado cierto de padecer el estado crónico i (p.ej., diálisis hospitalaria) durante toda la vida (t años). La probabilidad p se modifica hasta que el interrogado se muestra indiferente entre las dos alternativas, en cuyo caso el valor de preferencia del estado i se iguala a p ; esto es, $h_i = p$.

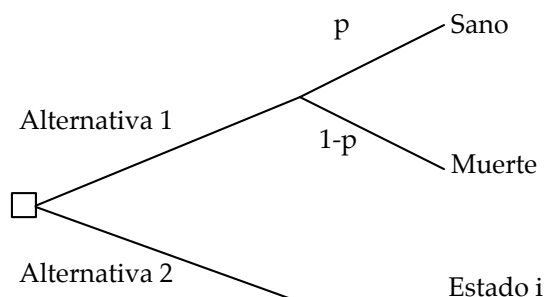


Figura. 1.6.- Jugada estándar para un estado de salud crónico preferido a la muerte.

La Figura 1.7 muestra la jugada estándar aplicada a estados peores que la muerte. La probabilidad p se modifica hasta que el sujeto se muestra indiferente entre las dos alternativas, en cuyo caso el valor de preferencias para el estado i es $h_i = -p / (1-p)$. Esto significa que los estados peores se representan por números negativos.

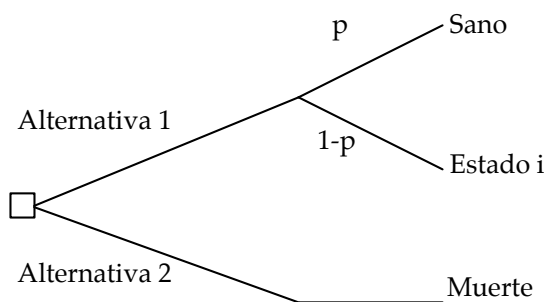


Fig. 1.7.- Jugada estándar para un estado de salud crónico considerado peor que la muerte.

La jugada estándar puede aplicarse a estados transitorios como se muestra en la Figura 1.8. Aquí la alternativa cierta es el estado i , el estado cuyo valor de preferencia debe medirse, igual que en el ejemplo del estado de salud crónico de la Figura 1.7. La diferencia entre las Figuras 1.6 y 1.8 es que el juego de la Figura 1.6 reemplaza la “muerte” por el peor estado transitorio (estado j). La fórmula que se usa para calcular el valor del estado i es: $h_i = p + (1-p)h_j$ (TORRANCE, 1976).

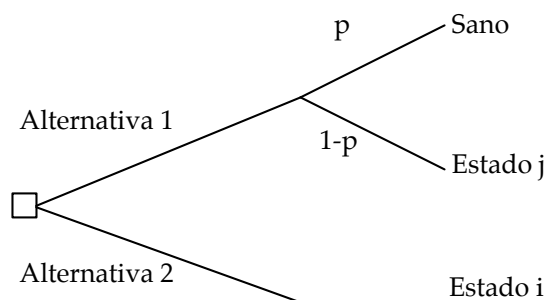


Fig. 1.8.- Jugada estándar para un estado de salud transitorio.

El **intercambio temporal** (*time trade-off*) fue desarrollado por TORRANCE y col. (1972) para usar específicamente en investigación de la salud como una alternativa a la jugada estándar más sencilla de administrar. En esta técnica se plantean dos alternativas al individuo encuestado: o vivir t años en el estado de salud i después de los cuales morirá o estar sano durante $X < t$ años después de los cuales morirá. Por tanto, con este procedimiento se valora la cantidad de vida (tiempo) frente a la Calidad de Vida. Se varía el tiempo X hasta que el encuestado manifiesta que le da igual una alternativa que otra. Entonces la preferencia para el estado i será: $h_i = X / t$.

El procedimiento anterior puede adaptarse para considerar estados de salud que son percibidos como peores que la muerte, así como para contemplar la transitoriedad de los estados.

La Figura 1.9 muestra el método del intercambio temporal para estados crónicos mejores que la muerte.

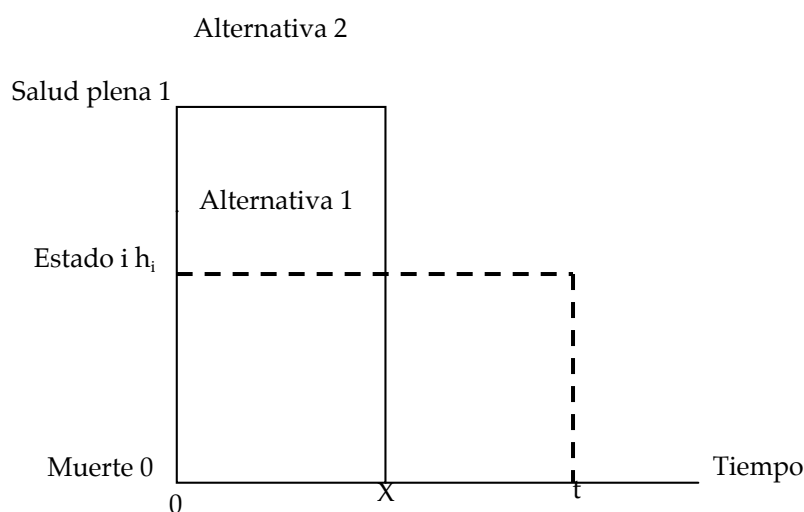


Figura. 1.9. - Intercambio temporal para un estado de salud crónico preferido a la muerte.

La **escala de puntuación** (*rating scale*) consiste en una línea con puntos finales o anclajes claramente definidos. Requiere que el interrogado identifique el mejor y el peor estado de salud para usarlos como anclajes. (En la práctica los anclajes se marcan generalmente como “muerte” y “salud perfecta”). El interrogado entonces puntúa la deseabilidad de cada estado de salud situándolo en algún punto de la línea entre los anclajes. Para conseguir una escala de intervalo, los interrogados deben recibir instrucciones para situar los estados de salud en la línea de tal forma que los intervalos entre los puntos representados reflejen las diferencias percibidas entre los estados de salud.

Si se considera la muerte como el peor estado de salud y la vida normal y sana como el mejor, se asigna a estos estados valores 0 y 1, respectivamente. Los demás estados de salud se puntúan en dicha escala. Pero si el peor estado no es la muerte sino otro estado situado en el punto p , la referencia a los distintos estados de salud se hace mediante el cociente $\frac{X - p}{1 - p}$ siendo X la ubicación del estado de salud en dicha escala.

Para evaluar las preferencias para estados de salud transitorios, se emplea un procedimiento similar. La única diferencia es que el mejor estado de salud se describiría como aquél sin enfermedad y se situaría en el 1.

La **estimación de la magnitud** es un método de escalamiento propuesto por STEVENS en 1971. Con este método, el paciente señala un estado de salud estándar y se le pide que proporcione un número o razón que indique cuánto mejor o peor es cada estado de salud comparado con el estándar.

Los estudios en los que se ha usado este método no han concordado en la selección de un estado de salud estándar. Estos estudios también difirieron en la dirección de la escala: algunos emplearon el 0 para definir el estado de salud menos deseable, mientras que otros lo utilizaron para el más deseable.

Otros dos métodos de escalamiento, el de equivalencia y el de la voluntad de pago, se han empleado menos frecuentemente pero son dignos de mención. El **método de equivalencia** ha sido aplicado en varias formas, aunque la tarea común subyacente para

el paciente es identificar el punto de indiferencia entre mantener vivo a un grupo de personas con un estado de salud A y mantener vivo a otro grupo mayor de personas, cuyo tamaño es indicado por el encuestado, con un estado de salud B peor que el estado de salud A. La relación entre el número de personas en cada grupo sirve para calcular las utilidades (en este caso, valor social o “utilidad social”). Esta técnica es conceptualmente similar a la estimación de la magnitud.

THOMPSON (1986) recomienda la técnica de **la voluntad de pago** para medir las preferencias de salud. Este método consiste en la formulación de la siguiente pregunta: ¿qué porcentaje de sus ingresos familiares estaría dispuesto a pagar para conseguir la curación completa de una determinada enfermedad?

Por último, para acabar este apartado, debemos mencionar que, como señala KAPLAN en 1997, hay desafíos tanto teóricos como metodológicos asociados con la aplicación de medidas basadas en la preferencia y la utilidad. A menudo, los pacientes pasan un rato difícil aplicando los métodos basados en la utilidad. La tarea ha sido facilitada por nuevos desarrollos en el software informático. Un ejemplo excelente es el programa U-Titer, desarrollado para el Macintosh.

1.2.5.- MODOS DE ADMINISTRACIÓN DE LOS CUESTIONARIOS

La herramienta básica y estándar de medida de la Calidad de Vida Relacionada con la Salud es el cuestionario. Los cuestionarios son cumplimentados por entrevistadores especializados o son autocumplimentados por el propio paciente. El primer método disminuye el número de preguntas no contestadas, mientras que el segundo es menos caro pero aumentan las contestaciones “perdidas”.

La recogida de la información puede realizarse por correo, lo que es propio de los cuestionarios autocumplimentados, por teléfono o personalmente; en estos dos últimos casos es el entrevistador el que registra las preguntas.

Los últimos desafíos metodológicos están orientados a mejorar la calidad de la medida de la CVRS mediante nuevas versiones de cuestionarios que **reduzcan el número de**

ítems y que además puedan ser administrados por ordenador (Computer Assisted Instruments, CAT) (PANE y col., 2006).

Una vez decidido cómo se va a realizar el cuestionario y el tipo de preguntas a formular se debe decidir sobre las codificaciones, puntuaciones y escalas de respuestas a utilizar. Esta fase es sumamente importante, pues una pregunta cerrada puede obviar resultados evidentes en el estudio y una pregunta abierta puede dificultar en exceso la evaluación de los resultados.

Medidas individualizadas

En los últimos años ha incrementado el interés en la investigación sobre el uso de medidas individualizadas que reflejan que la percepción de la Calidad de Vida es única para los individuos y no puede ser valorada adecuadamente usando medidas estándar.

Las medidas individualizadas más utilizadas son: El programa para la evaluación de la Calidad de Vida individualizada (Schedule for the Evaluation of Individual Quality of Life, SEIQOL); El índice generado por el paciente (Patient Generated Index, PGI, SKEVINGTON, 1999) y El Perfil de Repercusión de la Enfermedad (Disease Repercusión Prolife, DRP).

El uso de estas medidas ha sido limitado por dificultades en la administración. Primero, algunos pacientes tienen dificultad para comprender el sistema de la ponderación directa; segundo, los pacientes pueden no ofrecer voluntariamente algunos factores que son importantes para ellos, particularmente los relacionados con el humor (VACHON, y col. 1995), y la información que un individuo desea ofrecer voluntariamente cambia a lo largo del tiempo (HIGGINSON y col., 1994). Por último la interpretación y el análisis de algunos de los datos son complejos debido a su naturaleza individualizada. Se necesita más investigación para refinar y evaluar estas medidas individualizadas.

1.3. - PROPIEDADES PSICOMÉTRICAS DE LOS INSTRUMENTOS DE MEDIDA

1.3.1.- EVALUACIÓN CIENTÍFICA DE LOS INSTRUMENTOS DE MEDIDA DE LA CVRS

Los instrumentos para la medida de la Calidad de Vida Relacionada con la Salud deben ser evaluados en términos tanto de utilidad clínica como de solidez científica (HOBART y col., 1996). Para que un instrumento sea útil y aceptable clínicamente de tal manera que pueda ser incorporado a la práctica diaria, debe de ser adecuado al grupo de pacientes que está siendo estudiado, accesible a los usuarios, breve, fácil de administrar y costo-efectivo. Los instrumentos difíciles de manejar, que consumen tiempo y recursos, han limitado su uso en la práctica clínica. La utilidad clínica, no obstante, no garantiza la solidez científica en términos de medidas rigurosas.

La evaluación científica de los instrumentos de medida de la Calidad de Vida comprende la evaluación de la fiabilidad, validez y sensibilidad al cambio (NUNNALLY, 1978; CARMINES, y ZELLER, 1979; KAPLAN, y SACCUZZO, 1993; STREINER, y NORMAN, 1995).

1.3.1.1.- FIABILIDAD

Una medida fiable es aquélla que produce resultados que son exactos, consistentes, estables a lo largo del tiempo y reproducibles. Si estamos “midiendo” Calidad de Vida, se debe dar una estabilidad en los resultados cuando se repite el proceso de medición, en circunstancias similares.

Podemos diferenciar cuatro tipos diferentes de fiabilidad:

1ª) Consistencia interna (consistencia inter-ítem). Es el grado en que los ítems de una escala miden los mismos conceptos; esto es, la medida de homogeneidad de la escala. Se evalúa con el coeficiente alfa de Cronbach, en el caso de escalas de medida

continuas (CRONBACH, 1951), o con el estadístico 20 de Kuder-Richardson, en el caso de escalas de medida dicotómicas (KUDER y RICHARDSON, 1937).

El coeficiente *alfa de Cronbach*, es el más utilizado. Este coeficiente determina la consistencia interna de una escala analizando la correlación media de una variable con todas las demás que integran dicha escala. Toma valores entre 0 y 1, aunque también puede mostrar valores negativos (lo que indicaría que en la escala hay ítems que miden lo opuesto al resto). Cuando más se acerque a la unidad, mayor será la consistencia interna de los indicadores en la escala evaluada, aunque no existe un acuerdo generalizado sobre cual es el límite que demarca cuándo una escala puede ser considerada como fiable o no.

Según BERZON (2000) en general, es óptimo tener alfas por encima de 0.85, para proporcionar una estimación de una homogeneidad de dominio. Para valoraciones individuales de pacientes, se recomienda que los valores deberían ser superiores a 0.9. Un coeficiente alfa inferior podría indicar que el ítem no proviene del mismo dominio conceptual.

2ª) Fiabilidad test-retest. Evalúa la estabilidad de un instrumento de medida a lo largo del tiempo. Se evalúa aplicando la misma prueba dos veces al mismo grupo de sujetos con un intervalo de tiempo dado entre una y otra aplicación y examinando la correlación entre las dos puntuaciones. Suele utilizarse para instrumentos autoadministrados.

En investigación clínica es frecuente la evaluación de la fiabilidad a través de la comparación del acuerdo o desacuerdo producido en diferentes mediciones. Cuando la variable implicada es cuantitativa continua, suele utilizarse el coeficiente de correlación de Pearson (r), aunque su uso con esta finalidad no es muy correcto, ya que mide la intensidad de la asociación lineal entre las mediciones pero no proporciona información sobre el acuerdo observado al ignorar las diferencias sistemáticas ocurridas.

Una aproximación más adecuada es el coeficiente de correlación intraclase (ICC), que permite evaluar la concordancia general entre dos o más métodos de medida u observaciones diferentes. El coeficiente de correlación intraclase (ICC) se ha aceptado como el índice de concordancia para datos continuos. Hay diferentes versiones del ICC

que pueden dar resultados muy diferentes cuando están aplicados al mismo conjunto de datos (SHOUKRI, 2004).

La definición matemática del ICC se deriva de un modelo de análisis de la varianza de efectos mixtos. La varianza total entre las mediciones se debe a tres diferentes fuentes: las diferencias entre los observadores e instrumentos, las diferencias entre los pacientes y los residuos que representan la variación no explicada (KRAMER y FEINSTEIN, 1981).

Se define como la proporción de la variabilidad total que es debida a la variabilidad de los pacientes y se calcula a partir de las sumas de cuadrados del análisis de la varianza con medidas repetidas de la siguiente manera (BLAND, y ALTMAN, 1996):

$$ICC = \frac{k \cdot SC_{Entre} - STC}{(k-1)STC}$$

siendo k el número de medidas tomadas en cada sujeto.

En la siguiente tabla (Tabla 1.3) aparecen los valores del índice ICC y el grado de acuerdo con la escala, propuesta por LANDIS y KOCH (1977b):

Valor	Grado de acuerdo
0	Pobre
0.01-0.20	Leve
0.21-0.40	Regular
0.41-0.60	Moderado
0.61-0.80	Substancial
0.81-1.00	Casi Perfecto

Tabla 1.3.- Valores del índice ICC y grado de acuerdo con la escala según LANDIS y KOCH (1977b)

Entre las limitaciones de este coeficiente, podemos destacar las siguientes (PRIETO y col., 1998): es una prueba paramétrica, por lo que deben valorarse las posibles violaciones de los supuestos subyacentes; depende de la variabilidad de los valores observados, de forma que tiende a ser mayor en muestras heterogéneas y se expresa en

términos absolutos, por lo que debe tenerse en cuenta la significación clínica de las diferencias observadas.

3ª) Fiabilidad del evaluador. Es el acuerdo entre observadores o dentro de un observador individual. Es pertinente, por tanto, cuando el proceso de medida involucra un observador, pero no es relevante para instrumentos autoevaluados.

Distinguimos dos tipos de fiabilidad del evaluador: la fiabilidad inter-observador que es el acuerdo entre dos o más observadores y la fiabilidad intra-observador que es el acuerdo entre evaluaciones hechas por un único observador sobre el mismo paciente.

Para evaluar la fiabilidad entre observadores se utiliza el índice Kappa. El proceso de elaboración de este índice es la siguiente: se calcula la diferencia entre la proporción de acuerdo observado y la proporción de acuerdo esperado por azar; si ésta es igual a cero, entonces el grado de acuerdo que se ha observado puede atribuirse enteramente al azar; si la diferencia es positiva, ello indica que el grado de acuerdo es mayor que el que cabría esperar si solo estuviera operando el azar. Kappa es el cociente entre esa cantidad y el acuerdo máximo que se puede esperar sin intervención del azar.

$$Kappa = \frac{P_o - P_e}{1 - P_e}$$

donde P_o es la proporción de concordancia observada y P_e es la proporción de concordancia esperada.

El valor máximo de Kappa es uno si se da una total coincidencia entre los observadores. En la práctica se asumen los valores dados por FLEISS (1981): un Kappa de 0.8 o más, es necesario para que la escala sea fiable.

4ª) Fiabilidad de cuestionarios paralelos (fiabilidad de cuestionarios alternativos). Es el grado de acuerdo entre dos formas idénticamente construidas (semejantes) de la misma medida. Se evalúa preparando dos instrumentos de medida equivalentes, con la misma longitud, el mismo número de ítems, etc.; las dos formas se aplican al mismo grupo de sujetos, con o sin intervalo de tiempo, para obtener así la correlación de Pearson entre ambas aplicaciones.

1.3.1.2.- VALIDEZ

Aunque un instrumento debe ser fiable para que sea válido, la evidencia de fiabilidad no garantiza la evidencia para la validez; esto es, la fiabilidad es condición necesaria pero no suficiente, para la validez.

La validez puede definirse de un modo amplio como el grado en que el instrumento mide el concepto que se propone o está encaminado a medir. Para un instrumento que mida Calidad de Vida nos preguntaremos **¿El instrumento mide la Calidad de Vida del paciente?**

Hay tres tipos de validez:

1º) Validez de contenido: Es el grado en que una medida es representativa del dominio conceptual que está encaminada a cubrir. La validez de contenido constituye un elemento fundamental en toda prueba, pero no es suficiente para garantizar niveles superiores de validez. En estudios de Calidad de Vida nos preguntaríamos: **¿Los componente del cuestionario cubren TODOS los aspectos de la Calidad de Vida?.**

Ofrece dos modalidades:

Validez aparente (*face validity*). Evaluada por el juicio global de expertos sobre si el instrumento mide lo que se propone medir.

Validez lógica o muestral: Evaluación del grado en que el instrumento incluye todos los aspectos relevantes y representativos del dominio.

2º) Validez relacionada con el criterio. Es el grado en que una medida se correlaciona con un buen “patrón oro” (*gold standard*), otro método de medida, generalmente más complejo y costoso y para el que ya ha quedado demostrado que mide lo que pretende medir, que sirve de referencia o criterio. Para valorar la validez de criterio en Calidad de Vida, nos preguntaríamos: **¿El instrumento se comporta como el estándar, en la evaluación de la Calidad de Vida?.**

La validez es concurrente cuando ambas evaluaciones son simultáneas.

La validez es predictiva cuando las puntuaciones del criterio son obtenidas con posterioridad a la aplicación de la prueba y, por tanto, el instrumento de medida pretende predecir alguna conducta futura.

3º) Validez de “constructo”. Es el proceso utilizado para establecer la validez de un instrumento de medida a través de una serie de estudios que examinan la relación entre la medida y otras medidas o comportamientos. En ausencia de un “patrón oro” con el que comparar una medida, procede evaluar esta clase de validez. La estrategia fundamental consiste en comparar el instrumento de medida que queremos validar no sólo con los que miden el mismo atributo, sino también con otros que tratan contenidos diferentes. Este proceso se realiza en el contexto de un marco teórico en el que quedan explicitadas anticipadamente las diversas relaciones que podemos encontrar entre las puntuaciones obtenidas en la prueba y las obtenidas en las pruebas de contenidos similares y contenidos diferentes. Así, en la práctica, la evidencia para la validez de “constructo” se reúne emprendiendo una serie de estudios para determinar:

- El grado en que la medida se correlaciona con otras medidas de entidades relacionadas (validez convergente).
- El grado en el que la medida no se correlaciona con las medidas de diferentes entidades (validez discriminante).
- El grado en que la medida es capaz de detectar diferencias en grupos que se sabe difieren en el concepto que está siendo medido.
- El grado en que las hipótesis generadas sobre las bases de alguna noción teórica acerca del “constructo” son apoyadas por los resultados obtenidos de la medida.

La validez del “constructo” es, por tanto, lograda por la acumulación de todos estos tipos de evidencia. En vista de la falta de “patrones oro” en áreas tales como la evaluación de la discapacidad y minusvalía, la validez de “constructo” es la piedra angular de la evaluación de la validez de muchos instrumentos de medida de salud.

Si queremos evaluar la Validez de Constructo en un instrumento para evaluar el constructo Calidad de Vida nos preguntaríamos: **¿Los datos recogidos con el instrumento evidencian las dimensiones latentes esperadas?**.

Aunque hay algún solapamiento entre los tres tipos de validez, cada una tiene un enfoque algo diferente para evaluar el grado en que un instrumento mide lo que se propone.

Por tanto, los tres tipos de validez deben considerarse para establecer un juicio global de la exactitud de las inferencias extraídas de un test (MESSICK, 1980).

COX y col., (1992), recomiendan usar instrumentos estándar validados, a los que se les puede añadir ítems específicos para una aplicación particular. Estos autores señalan que la facilidad en la aplicación del cuestionario a los individuos es fundamental, no sólo cuando se pasa a pacientes sino también cuando se pasa a individuos sanos.

Como técnica estadística para contrastar la validez del constructo se usa el Análisis Factorial Exploratorio (EFA). Esta técnica permite resolver el problema del análisis de la estructura de las interrelaciones (correlaciones) existentes en un número elevado de variables (que en este caso serían los ítems del cuestionario) definiendo un número de dimensiones comunes subyacentes, denominadas factores.

El **Análisis Factorial** proporciona un procedimiento para contrastar la verdadera dimensionalidad de las escalas. Al aplicar el Análisis Factorial se pondrá de manifiesto si el constructo evaluado tiene una dimensión o no; es decir, si es realmente multidimensional como solemos hipotizar a priori.

El modelo del Análisis Factorial viene dado por las ecuaciones:

$$\begin{aligned} X_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + U_1 \\ X_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + U_2 \\ &\vdots \\ X_n &= a_{n1}F_1 + a_{n2}F_2 + \dots + a_{nm}F_m + U_n \end{aligned}$$

donde F_1, F_2, \dots, F_m son los factores comunes con ($m < n$) y U_1, U_2, \dots, U_n los factores únicos o específicos y los coeficientes a_{ij} las cargas factoriales y h_i^2 la communalidad de $X_i = \sum a_{ij}^2$. Se supone, que los factores comunes están a su vez estandarizados y tanto los factores comunes y específicos están incorrelados.

Para obtener la matriz $A = (a_{ij})$, existen diversas técnicas entre las que destacamos: Método del factor principal, Análisis de Componentes Principales y el Método de Máxima Verosimilitud. A continuación describimos el análisis de Componentes principales ya que es el más utilizado en estudios de validación de cuestionarios de medida de Calidad de Vida Relacionada con la Salud.

Este método de estimación de la matriz factorial considera sólo los factores comunes y prescinde de los factores únicos. El procedimiento se reduce entonces a calcular los valores y vectores propios de la matriz de correlaciones calculada a partir de los datos.

Por lo tanto, es una técnica estadística que permite transformar un conjunto de variables intercorrelacionadas en otro conjunto de variables no correlacionadas, denominadas factores o componentes principales.

Números de factores a retener

En la práctica una de las decisiones principales es el número de factores a retener para obtener una explicación suficientemente buena de las correlaciones o covarianzas entre las variables observadas.

Se han establecido diversas reglas y criterios para determinar el número de factores a conservar, de manera que se pueda cumplir el principio de parsimonia². Algunos de los más frecuentes los señalamos a continuación.

² Principio de Parsimonia: Es el criterio según el cual debe considerarse como válida, entre varias soluciones admisibles, aquella que contenga el número mínimo de factores comunes ($m < n/2$).

- Regla de Kaiser. Sugiere conservar sólo aquellas componentes principales cuyos valores propios sean mayores que uno.
- Regla del 75% de la varianza. El número de factores será suficiente cuando el porcentaje de varianza total acumulada sea de un 75%.
- Scree Plot o regla del codo (CATTELL y JASPERS, 1967). Regla empírica que se basa en la utilización de una representación gráfica donde los factores están en el eje de abscisas y los valores propios en el de ordenadas. El punto de distinción vienen representado por el punto de inflexión de la gráfica. Se pueden conservar los factores situados antes de este punto.
- Método por defecto. Retiene el número más elevado de entre los posibles que proporcionan los criterios de selección habituales.

Rotación de factores

Es muy frecuente en estudio psicométricos que la solución encontrada en el Análisis Factorial no sea fácilmente interpretable. Los investigadores generalmente someten la solución a una rotación para conseguir otra que sea más interpretable. En los trabajos de validez del constructo, en estudios de Calidad de Vida, la única alternativa encontrada aplicada es la rotación ortogonal Varimax.

Se pretende transformar los ejes de modo que los nuevos factores sigan siendo incorrelados:

$$\mathbf{B} = \mathbf{AT}$$

siendo **A** la matriz factorial inicial, **T** la matriz de una transformación ortogonal y **B** la matriz factorial final.

La rotación ortogonal Varimax maximiza la suma de varianzas de las cargas requeridas ($\text{Var}(b_{1j}^2, \dots, b_{nj}^2)$) de la matriz de factores. La rotación Varimax es la que permite obtener las cargas más extremas (cercanas a -1 o al +1) y otras cargas cercanas al 0. El interés de esta rotación es que permite interpretar los factores más fácilmente, al indicar una asociación positiva o negativa clara entre la variable (ítem) y el factor (o una

ausencia de asociación si el valor está cercano a 0). Mostrándose, pues, la forma más clara de separar los factores.

1.3.1.3.- SENSIBILIDAD AL CAMBIO (*responsiveness*)

Aunque la fiabilidad y la validez son los principales determinantes de la solidez científica de una medida, la capacidad de un instrumento para detectar un cambio clínicamente significativo es también esencial cuando se evalúan los beneficios relativos de diferentes intervenciones. Esto es particularmente importante cuando los tratamientos están asociados con diferencias pequeñas pero significativas (una característica de las intervenciones en la actualidad) que no pueden ser detectadas por medidas que son insensibles al cambio. En tales casos un instrumento apropiado, fiable y válido pero insensible al cambio es de valor limitado.

La sensibilidad al cambio de un instrumento puede ser determinada mediante los siguientes procedimientos (DEYO y col., 1991; GUYATT y col., 1987):

- La administración consecutiva de la medida en diferentes puntos a lo largo del tiempo cuando se espera que ocurran cambios clínicos (por ejemplo, antes y después de un tratamiento de conocida eficacia).
- La comparación frente a otros criterios de cambio (por ejemplo, percepciones de cambio del personal y del paciente).
- La comparación con otras medidas del mismo concepto para evaluar la sensibilidad al cambio relativa.

Diversos autores (ALLISON y col., 1997; SPRANGERS y SCHWARTZ, 1999) postulan que la Calidad de Vida Relacionada con la Salud es un “constructo” dinámico que cambia en respuesta a ciertos procesos como la capacidad de adaptación, la capacidad de hacer frente a un estado de salud, las experiencias vitales vividas, las expectativas futuras, etc. Así pues, para comprobar la eficacia de un determinado tratamiento en términos de la Calidad de Vida Relacionada con la Salud no sólo interesa emplear un instrumento de medida sensible al cambio que detecte si éste se produce cuando realmente ocurre, sino que también interesa separar las fuentes del cambio

detectado en cambio debido al tratamiento y cambio al dinamismo del “constructo”. Procediendo de esta forma, podremos determinar si el cambio observado se debe al tratamiento aplicado o bien puede explicarse por los procesos a los que antes hacíamos referencia. Esta cuestión es objeto de estudio en la actualidad.

Entre los estadísticos para valorar la sensibilidad al cambio de los cuestionarios destacamos:

El tamaño del efecto (ES). (KAZIS y col., 1989). Es el cambio medio en las puntuaciones, dividido por la desviación estándar de las puntuaciones de la Calidad de Vida obtenidas en la primera ocasión. Es muy útil cuando se necesita comparar resultados de diferentes estudios o de diferentes instrumentos, ya que los resultados no tienen unidades de medida.

La respuesta media estandarizada (SRM) (COHEN, 1977; LIANG y col., 1990). Compara la magnitud del cambio con la desviación estándar del cambio. Un valor grande de SRM indica alta sensibilidad al cambio.

El estadístico de responsiveness de GUYATT (RS). Compara la magnitud del cambio con la desviación estándar del cambio entre sujetos estables; es decir, aquellos cuyo verdadero estatus es constante a lo largo del tiempo.

1.3.2. - PROPIEDADES PSICOMÉTRICAS DE LOS INSTRUMENTOS DE MEDIDA DE CVRS MÁS UTILIZADOS

Perfil de las Consecuencias de la Enfermedad (*Sickness Impact Profile*, SIP)

El SIP fue traducido y adaptado al español con el nombre de Perfil de las Consecuencias de la Enfermedad (BADÍA y ALONSO, 1994). Con la versión final se realizó un estudio piloto en 50 pacientes en los que se comprobó que la comprensión de los ítems era buena.

Se analizó la validez discriminante comparando las puntuaciones medias del Perfil de las Consecuencias de la Enfermedad de diferentes subgrupos de pacientes y sanos, que

mostró que la diferencia entre las puntuaciones medias era estadísticamente significativa entre todos los subgrupos salvo en dos (BADÍA y ALONSO, 1995).

Como en la versión original, se analizó la validez convergente del Perfil de las Consecuencias de la Enfermedad comparando sus puntuaciones con otras variables de criterio. Se obtuvieron correlaciones bajas con las medidas clínicas de disfunción y enfermedad, sugiriendo que el Perfil de las Consecuencias de la Enfermedad aporta una información no registrada sistemáticamente por los clínicos.

La consistencia interna, medida mediante el coeficiente alfa de Cronbach, para la puntuación total del Perfil fue alta, siendo moderada a baja para las categorías de nutrición y sueño-descanso y moderada a alta para el resto.

La fiabilidad test-retest se midió en una muestra de 129 pacientes, obteniéndose una correlación alta en todas las categorías y para el total. La fiabilidad se mantuvo alta y estadísticamente significativa para cada subgrupo de pacientes, modo de administración (entrevista versus autoadministrado) y para cada diferente entrevistador (BADÍA y ALONSO, 1995).

La sensibilidad al cambio de la versión española no ha sido evaluada aunque existen estudios en marcha en el ámbito de la reumatología y la insuficiencia renal crónica (BADÍA y col., 1999).

Cuestionario de salud SF-36 (SF-36 *Health Survey*)

La traducción al español del cuestionario siguió un protocolo común en todos los quince países que participan en un proyecto internacional de adaptación del cuestionario original (AARONSON y col., 1992), basado en el método de traducción y retrotraducción por profesionales y la realización de estudios piloto con pacientes para valorar la comprensión y la factibilidad de administración del cuestionario.

La validez ha sido evaluada en un estudio llevado a cabo con 321 pacientes (hombres) con enfermedad pulmonar obstructiva crónica, confirmando la validez de concepto de la adaptación.

Se comprobó la validez de “constructo” del SF-36 mediante un análisis de componentes principales, en el que se obtuvieron dos componentes (física y mental) que explicaron el 65,6% de la variabilidad total. Este patrón factorial demostró ser estable en diferentes países y en submuestras basadas en grupos de edad y sexo (WARE y col., 1998).

El SF-36 se administró a una muestra de 9151 individuos residentes en el territorio español. La consistencia interna, estimada mediante el coeficiente alfa de Cronbach, se situó en todos los casos por encima del valor 0.7 (valor mínimo recomendado).

También se calculó el coeficiente alfa de Cronbach para la muestra de 321 hombres con enfermedad pulmonar obstructiva crónica, comprobándose así mismo un alto nivel de consistencia interna para todas las dimensiones exceptuando la función social (ALONSO y col., 1994).

Perfil de Salud de Nottingham (*Nottingham Health Profile, NHP*)

La adaptación del NHP al español se inició en 1985 y siguió el procedimiento utilizado previamente para el desarrollo de las versiones sueca y francesa del cuestionario. La versión final fue probada en un estudio piloto en el que se demostró fácil de entender y de completar (ALONSO y col., 1990; ALONSO y col., 1994).

A partir de una base de datos común que incluyó 45 estudios diferentes y variables como el diseño del estudio, la edad, el sexo, el nivel socioeconómico, la localidad, el problema de salud principal, la comorbilidad, etc., se analizó la calidad de la información (datos incompletos, la consistencia interna y la distribución de las puntuaciones del cuestionario), la fiabilidad y la validez del instrumento y los factores que influyen sobre estas características.

Los resultados confirmaron la adecuación de las propiedades métricas de la versión original del Perfil de Salud de Nottingham. Fiabilidad y validez fueron evaluadas en diferentes grupos poblacionales, obteniéndose resultados similares en todos ellos. En general, la proporción de datos faltantes fue baja, observándose sin embargo un “efecto techo” de las puntuaciones del Perfil de Salud de Nottingham moderado-alto.

La sensibilidad de la versión española del cuestionario fue evaluada con 104 mujeres embarazadas que contestaron al cuestionario a las 18, 27 y 37 semanas de gestación, mostrándose que las puntuaciones incrementaron su valor para la mayoría de las dimensiones a medida que avanzaba la gestación (BADÍA y col., 1999).

EuroQol (EQ-5D)

En 1992, se adaptó el instrumento para su uso en España. Se realizó una prueba piloto con 10 pacientes y 10 individuos aparentemente sanos que sugirió algún cambio y mostró que la traducción era satisfactoria.

Se evaluó la validez de concepto mediante un análisis factorial que mostró tres factores: “físico”, “dolor” y “salud mental” (RUÉ y BADÍA, 1996). Los resultados muestran que, en todos los casos, el EuroQol es capaz de discriminar entre mala salud y buena salud según las hipótesis previamente establecidas (BADÍA y col., 1996).

Se realizó un estudio administrando el cuestionario en dos ocasiones separadas por un intervalo de 15 días, para comprobar la fiabilidad test-retest obteniendo coeficientes de correlación de Spearman altos, moderados y bajos en distintos estados de salud (ALEGRE y NUALART, 1995).

Se ha estudiado recientemente el cambio en la Calidad de Vida antes y después del tratamiento en pacientes con esquizofrenia; y antes y después del tratamiento antiretroviral, obteniéndose un tamaño del efecto de 1.2 y 0.6, respectivamente en la EVA y el índice de EuroQol.

Cuestionarios QUALEFFO y OQLQ

El Cuestionario QUALEFFO fue validado en un estudio multicéntrico llevado a cabo en 7 países (LIPS y col., 1999) sobre pacientes de 55-80 años con osteoporosis clínica. Los controles se estratificaron por edad y sexo y no presentaban dolor crónico ni fracturas vertebrales. Se administró el QUALEFFO 2 veces en 4 semanas y se comparó con el SF-36. La fiabilidad test-retest fue buena (0,54-0,90) y la consistencia interna de cada una de las 5 dimensiones fue acta (alfa de Cronbach, 0,80). Las puntuaciones

medias del instrumento fueron significativamente más altas en pacientes con fracturas vertebrales que en los controles en todas las dimensiones. Se encontraron correlaciones significativas entre dimensiones equivalentes del QUALEFFO y del SF-36, especialmente para el dolor, la función física y la función mental. Los *odds ratio* para el dolor y la función social fueron mayores para el QUALEFFO, mientras que la percepción general de salud fue más discriminativa para el SF-36.

El grupo de trabajo que desarrolló el cuestionario OQLQ, lo aplicó en 226 mujeres de Estados Unidos y Canadá (Osteoporosis Quality of Life Study Group, 1997) halla evidencias acerca de la excelente función discriminativa (fiabilidad en las 5 dimensiones entre 0.75 y 0.85 y altas correlaciones con medidas independientes de dolor, función física y estado emocional) y evaluativa (muy buena sensibilidad a los cambios clínicos producidos). Se plantean algunas dudas acerca de su validez longitudinal, basadas fundamentalmente en las pobres correlaciones entre los cambios producidos con el nuevo cuestionario y los detectados por el perfil de consecuencias de la enfermedad, aunque con el SF-36 las correlaciones sí se encuentran en el rango anunciado.

BADIA y col., (2001), realizaron un estudio de validación de los cuestionarios QUALEFFO y OQLQ haciendo una comparación entre ellos. El estudio fue realizado con una muestra de 338 mujeres diagnosticadas de osteoporosis y con deformidad vertebral y una muestra de 308 mujeres de una población normal (grupo control). Los resultados obtenidos fueron: para el OQLQ, un α de Cronbach de 0.75-0.91 y un ICC: 0.85-0.93, y para el cuestionario QUALEFFO, un α de Cronbach 0.63-0.90 y ICC: 0.80-0.93.

1.3.3.- PROCEDIMIENTOS PARA LA OBTENCIÓN DE MEDIDAS BASADAS EN UTILIDAD

La Tabla 1.4 expone los datos disponibles sobre cada tipo de fiabilidad para los diferentes procedimientos para valorar utilidades. Globalmente, estos datos son alentadores, aunque los huecos de la tabla indican la necesidad de investigación adicional. También, las comparaciones entre los estudios están limitadas por el hecho de que un estadístico frecuentemente usado, el coeficiente de correlación de Pearson, es

dependiente de la variabilidad a través de los sujetos. De este modo, las correlaciones de estudios que usan diferentes sujetos y tamaños muestrales no son directamente comparables (READ y col., 1984).

<i>Fiabilidad</i>	<i>JE</i>	<i>IT</i>	<i>EP</i>	<i>EM</i>	<i>EQ</i>	<i>VP</i>
Fiabilidad intra-evaluador	0.77	0.77-0.88	0.70-0.94	0.74-0.83		
Concordancia intra-evaluador (%)				97.2%		
Fiabilidad test-retest	0.80	0.87	0.77			
4 semanas		0.81				
6 semanas		0.63-0.80				
1 año	0.53	0.62	0.49			0.25
Fiabilidad Inter-evaluador			0.75-0.77	0.75-0.79		0.60
Concordancia Inter-evaluador (%)				88%		

Tabla 1.4.- Fiabilidad de los métodos de obtención de utilidades tomada de FROBERG y KANE (1989) donde: JE= Jugada estándar; IT= Intercambio temporal; EP= Escala de puntuación; EM= Estimación de la magnitud; EQ= Equivalencia; VP= Voluntad de pago.

Esta tabla no incluye las medidas test-retest realizadas antes y después de la terapia, o antes y después de algún otro suceso que probablemente cambiase las preferencias. A este respecto, los hallazgos discrepantes recogidos en la bibliografía (CHRISTENSEN-SZALANSKI, 1984; O'CONNOR y col., 1987; LLEWELLYN-THOMAS y col., 1984) señalan la necesidad de examinar las causas de los cambios en las preferencias. Por otra parte, son posibles muchos enfoques para comprobar la validez, dos de los cuales se han usado en la validación de los métodos de escalamiento de las preferencias de salud:

1.- *Examinar el grado de convergencia de los resultados de diferentes métodos de escalamiento.*

Los estudios que comparan métodos de escalamiento o han examinado las relaciones funcionales entre los métodos o han comparado los valores de escala medios obtenidos a partir de cada uno de los métodos.

La Tabla 1.5 resume los estudios que han comparado los resultados de diferentes métodos de escalamiento. Un “sí” indica que los investigadores hallaron al menos una de las tres condiciones:

1ª.- Una relación lineal entre los métodos de escalamiento.

2ª.- Una correlación significativa entre los métodos de escalamiento (lo que no necesariamente implica una relación lineal estricta).

3ª.- Que los valores medios no fueron estadísticamente diferentes.

<i>Estudio</i>	<i>JE</i>	<i>IT</i>	<i>EP</i>	<i>EM</i>	<i>EQ</i>	<i>V</i>
Patrick <i>et al.</i> (1973)			Sí	Sí	No	
Kaplan <i>et al.</i> (1979)			No	No		
Haig <i>et al.</i> (1986)			Sí	Sí		
Torrance (1976)	Criterio	Sí	No			
Wolfson <i>et al.</i> (1982)	No	Sí	Sí			
Read <i>et al.</i> (1984)	Sí	Sí	Sí			
Llewellyn-Thomas <i>et al.</i> (1984)	No		No			
O'Connor <i>et al.</i> (1987)	No		No			
Detsky <i>et al.</i> (1986)	Sí	No				
Churchill <i>et al.</i> (1987)	No	No				
Miles (1977)			Sí		Sí	

Tabla 1.5. - Convergencia de métodos de escalamiento tomada de FROBERG y KANE (1989) donde: JE= Jugada estándar; IT= Intercambio temporal; EP= Escala de puntuación; EM= Estimación de la magnitud; EQ= Equivalencia; VP= Voluntad de pago.

La tabla muestra que estos estudios han dado lugar a resultados variados.

Se necesita más investigación para examinar la convergencia de los métodos de escalamiento, especialmente los dos que no han sido aún estudiados, el de equivalencia y el de la voluntad de pago. Sin embargo, en la bibliografía de medición psicosocial se acepta generalmente que, aunque los diferentes métodos de escalamiento producen la misma ordenación según el rango, no se espera que den lugar necesariamente a resultados idénticos. Los valores de escala exactos obtenidos por los diferentes métodos diferirán debido a que los métodos piden a los encuestados que realicen tareas diferentes, recurriendo quizás a procesos cognitivos totalmente diferentes.

La selección de un método de escalamiento apropiado depende de la manera en que se usarán los resultados. Además, investigación adicional que aclare las relaciones entre los resultados de diferentes métodos de escalamiento y otros criterios externos

aumentará nuestra comprensión sobre lo que miden realmente estos métodos de escalamiento.

2.- Examinar el grado de apoyo empírico de las relaciones predichas y otras variables.

Pocos estudios han contrastado las relaciones hipotéticas entre las preferencias de los estados de salud y otras variables (CHURCHILL y col., 1987; KIND y col., 1982; THOMPSON y col., 1984; CHRISTENSEN-SZALANSKI, 1984).

1.4.- LIMITACIONES QUE SE PRESENTAN EN EL USO DE LAS MEDIDAS DE CALIDAD DE VIDA

A pesar del reconocimiento creciente de la utilidad de las medidas de Calidad de Vida, hay una diferencia notable entre sus posibles usos y su uso real. Esto se debe a que la aplicación de estas medidas encuentra cuatro tipos de obstáculos (DEYO y PATRICK, 1989) que resultan difíciles de superar. Éstos son:

1º) Obstáculos de actitud y de conocimiento

Las personas que tienen que aplicar las medidas de Calidad de Vida Relacionada con la Salud (médicos y personal sanitario) generalmente tienen poca o ninguna formación en la metodología de los cuestionarios y de las medidas de la utilidad. Por otra parte, consideran la información subjetiva obtenida directamente de los pacientes como poco “científica”, en contraposición con la información proporcionada por los datos fisiológicos o las observaciones físicas, que son inherentemente más exactos y reproducibles.

2º) Obstáculos conceptuales

Un importante problema conceptual en la aplicación de las medidas del estado de salud es la dificultad para distinguir entre los diversos propósitos de un cuestionario. KIRSHNER y GUYATT en 1985 han señalado que los cuestionarios tienen al menos tres objetivos: Discriminar entre los sujetos en un punto en el tiempo; predecir algún

resultado futuro o los resultados de un *gold standard* que implique más riesgos o que resulte más caro, y evaluar los cambios a lo largo del tiempo.

Estos autores sostienen que los requisitos de un instrumento dependen del propósito de éste y que, incluso, podría ocurrir que tales requisitos entrasen en conflicto uno con otro. Por ejemplo, un cuestionario usado para discriminar debería ser exhaustivo y preguntar acerca de los rasgos relativamente estables que distinguen a las personas en un punto del tiempo. Por el contrario, un instrumento usado con propósitos de evaluación podría no beneficiarse de la inclusión de características inmutables; resultaría más eficiente desde el punto de vista teórico si indagara únicamente sobre conductas y capacidades que son susceptibles de cambio como respuesta a la intervención en estudio.

3°) Obstáculos metodológicos

Muchos posibles usuarios no están familiarizados con la literatura sobre la valoración del estado de salud, incluso en su especialidad. Encontrar la literatura pertinente es muchas veces difícil, debido a las numerosas denominaciones empleadas para designar los instrumentos de medida del estado de salud, la escasez de términos de búsqueda en bases de datos informatizadas y la dispersión de artículos importantes en múltiples disciplinas.

Finalmente, otra barrera metodológica es que se desconoce si los instrumentos de medida del estado de salud funcionan igualmente bien en todos los marcos de aplicación. Tampoco está claro si un instrumento que resulta válido y aplicable en una cultura o en un grupo clínico será igualmente válido en otras culturas o en otros grupos. Así, a menudo es necesario evaluar de nuevo la validez y la fiabilidad y modificar las medidas cuando se usan en nuevos ambientes, proceso que resulta largo y costoso.

4°) Barreras prácticas

Quizá la barrera práctica más obvia es la carga excesiva, en tiempo y esfuerzo, que supone para la persona encuestada la cumplimentación de cuestionarios extensos.

Los costes derivados de la recogida y manipulación de los datos son otras barreras prácticas que provocan que la valoración del estado funcional sea relativamente más pesada y cara que los análisis de sangre o la obtención de datos a partir de registros médicos y bases de datos informatizadas.

Otro coste adicional es el de los estudios piloto que habitualmente es necesario realizar en nuevas aplicaciones, con el fin de comprobar el tiempo de administración del instrumento y el grado de aceptación del mismo. Los estudios piloto permiten, también, entrenar de forma adecuada a los entrevistadores que se encargarán de la recogida de datos.

Todos estos costes incrementan considerablemente el presupuesto de la investigación y dificultan la consecución de fuentes de financiación.

Finalmente, otro problema en todas las aplicaciones de la medida del estado de salud es la presentación de los resultados de una manera comprensible. Como se comentó anteriormente, tanto los médicos como los gestores no están familiarizados con estos instrumentos y las puntuaciones y los cambios en éstas deben acompañarse de alguna indicación que facilite su interpretación.