

**Mejora de la calidad de la gestión de datos clínicos a  
través de sistemas computerizados.**  
Aplicación a la informatización de la entrevista  
diagnóstica estructurada DICA-IV

Rosario GRANERO PÉREZ

Tesis doctoral dirigida por los Doctores:

Josep M. Doménech i Massons  
Lourdes Ezpeleta Ascaso

Departament de Psicobiologia i Metodologia de les Ciències de la Salut  
Facultat de Psicologia  
Universitat Autònoma de Barcelona

Este trabajo ha sido posible gracias a las ayudas DGICYT PM91-0209 y PM95-126 del  
Ministerio de Educación y Cultura

Bellaterra, 1998

## EL ORDENADOR COMO HERRAMIENTA DE AYUDA PARA MEJORAR LA CALIDAD DE LA GESTIÓN DE DATOS CLÍNICOS

*"The cost of trying to achieve error-free estimates is too high for most research purposes; some potential for error exists in virtually all survey plans. Total survey design involves considering all aspects of a survey and choosing a level of rigor appropriate to the particular purposes".*

(Fowler, 1993)

### 3.1. INTRODUCCIÓN

A pesar de que muchos clínicos e investigadores conocen la existencia y las ventajas que supone el uso de sistemas informáticos complejos como herramientas de ayuda para el diagnóstico y la toma de decisiones clínicas (por ejemplo, los sistemas expertos o las redes neuronales), ya hemos indicado con anterioridad que la incorporación de estos sistemas continúa siendo más una curiosidad experimental y empírica que una realidad práctica (Shanteau, 1992a, 1992b). El por qué de esta actitud resulta, cuanto menos, evidente. Además de las razones habitualmente argumentadas referidas a la falta de conocimiento de gran parte del "software" disponible a nivel comercial, si tenemos en cuenta que habitualmente la selección de las pruebas de evaluación y de las herramientas de ayuda clínica es en sí una tarea compleja de toma de decisiones, entenderemos que la elección final depende de la valoración de diferentes criterios relacionados tanto con los objetivos profesionales como con las posibilidades y recursos reales disponibles para satisfacer dichos objetivos (Taft, 1993). Tal vez por ello, hoy en día la mayoría de clínicos e investigadores que trabajan en el ámbito del psicodiagnóstico (al menos en nuestro país) utilizan programas más sencillos que han sido diseñados para ser empleados, fundamentalmente, en las fases de recogida (captura), corrección y, sobre todo, en el análisis estadístico de las pruebas de evaluación psicológica habituales. La incorporación de este tipo de programas ha tenido como objetivos básicos disminuir el tiempo requerido para la evaluación psicológica, reducir los costes de la investigación, implementar diseños más complejos, realizar análisis más potentes, y paliar los frecuentes errores humanos detectados en las tareas cotidianas de registro, combinación de la información y análisis de datos. En relación con este último objetivo, se considera que, al menos de forma indirecta, una de las pretensiones básicas ha consistido precisamente en mejorar la calidad de la gestión de los datos que manejan los profesionales (Yeoh y Davies, 1993).

Paralelamente al uso de las nuevas tecnologías informáticas, durante los últimos años han aparecido diversos trabajos que revisan aspectos relacionados con la calidad de los datos y con las consecuencias que conlleva la falta de controles rutinarios de calidad (Magnusson y Bergman, 1990). Este interés se ha justificado argumentando que aunque la producción deliberada de errores en investigación científica es poco frecuente, el manejo deficiente de la información disponible podría dificultar tanto la verificación de la integridad de los datos clínicos como la reproducción de los análisis estadísticos y de los resultados en posteriores trabajos (Freedland y Carney, 1992). La posibilidad de que las deficiencias debidas a errores de gestión puedan aparecer en estudios dirigidos por investigadores responsables y altamente

competentes, ha dirigido numerosos esfuerzos para acabar con dichas fuentes de error, y también para orientar a los investigadores a la hora de realizar su labor con la máxima eficiencia en cuestiones de calidad (Grisso et al., 1991).

En relación con las líneas de investigación apuntadas en el último párrafo, diversos estudios se han esforzado en demostrar que el uso de programas de ordenador (tanto simples como complejos) no sólo conlleva una reducción sistemática en los costes de la investigación, la posibilidad de implementar diseños más complejos o la oportunidad de realizar análisis estadísticos más potentes, sino también una notable mejora en la calidad de los diagnósticos (Severe et al., 1989; Shanks, 1989). Otros trabajos, sin embargo, han matizado que las ventajas que se evidencian en la calidad de los resultados podrían ser debidas no tan sólo al uso de sistemas informatizados, sino a diversos factores relacionados con dicho uso: la definición clara de las variables a evaluar, la forma de registrarlas, y el hecho de que el clínico decida revisar y contrastar sus propios diagnósticos cuando éstos no concuerdan con los que le facilita el ordenador (Godoy, 1996). Finalmente, otras investigaciones también confirman que con el mero uso de instrumentos informatizados no siempre se consigue una mejora importante en la calidad de los juicios clínicos, ya que para conseguir incrementar la calidad de los resultados es imprescindible mejorar simultáneamente los procedimientos de recogida de datos, los algoritmos para generar los indicadores de adaptación psicológica, los métodos de análisis de datos y las estrategias para interpretar la información (Butcher, 1994; Cobos, 1995; Freedland y Carney, 1992; Siegel y Parrino, 1988). De hecho, aunque pueda parecer paradójico, en muchas ocasiones la falta de conocimiento de los procedimientos que se consiguen automatizar con las nuevas aplicaciones informáticas, unido a la confianza excesiva en que los resultados son correctos por el mero hecho de que se obtienen con un ordenador, influyen negativamente en la gestión de la información que se maneja y se publica.

Dentro de esta última línea de trabajo, algunos profesionales se lamentan de que muchos clínicos e investigadores, que se ven especialmente afectados por un síndrome al que se ha venido denominando como GIGO, iniciales de "Garbage in - Garbage out" (Cobos, 1995), olvidan la extrema importancia de comprobar la calidad del conjunto de operaciones realizadas en la difícil etapa del proceso de datos, introduciendo sesgos que afectan tanto a la validez interna como a la validez externa del estudio, y que pueden comprometer seriamente la veracidad de las conclusiones. En estos casos, ni el ordenador, ni el sistema de análisis estadístico más completo y sofisticado del mercado, pueden ayudar al investigador: cuando lo que se analiza es "basura", lo que se interpreta es, en el mejor de los casos, "basura con maquillaje". El efecto GIGO afecta especialmente las investigaciones que manejan grandes volúmenes de datos y/o cuándo éstos se estructuran de forma compleja, ya que en estas situaciones pueden aparecer serias dificultades para controlar el correcto funcionamiento de todas las operaciones que conlleva el manejo de la información (Barton et al., 1991; Moritz et al., 1995).

En efecto, las espectaculares ventajas que ofrecen las herramientas informáticas favorecen que el investigador se olvide que a lo largo del complejo proceso que integra un trabajo de investigación existen muchas oportunidades para producir errores en los datos. El propio proceso de observación o medición, la transcripción de los datos (al escribirlos sobre papel o al grabarlos en el ordenador) y la posterior manipulación (transformación y generación de nuevas variables) constituyen fuentes de error habituales en el proceso de gestión de datos. Esto implica que cuando se manejan grandes volúmenes de información, la presencia de errores no es una posibilidad remota o indeseable, sino prácticamente una certeza.

En un intento por sistematizar las deficiencias más habituales relacionadas con la baja calidad de los datos, algunos autores han querido destacar las principales fuentes de error (Gassman, Owen, Kuntz, Martin y Amoroso, 1995): (1) los errores de medida relacionados con las desviaciones o las interpretaciones erróneas de los protocolos; (2) el exceso de datos registrados para responder los objetivos de un estudio; (3) la documentación inadecuada de la

base de datos y/o del programa de análisis estadístico; (4) la existencia de varias bases de datos en un mismo estudio; (5) el uso inadecuado de las copias de las bases de datos; (6) el registro inadecuado de los datos en formato electrónico; (7) la dificultad para obtener indicadores de adaptación psicológica debido a factores tales como la presencia de valores desconocidos o la pérdida de cuestionarios; (8) la combinación inadecuada de la información, o lo que es igual, el manejo inadecuado de los datos en las fases de creación y modificación de variables; (9) el uso de técnicas estadísticas inadecuadas para estudiar los correlatos entre los indicadores psicopatológicos y los factores biopsicosociales; y (10) el entrenamiento inadecuado del personal que llevará a término la investigación. A su vez, estas deficiencias acontecidas en las fases de observación y gestión de datos conlleva unas importantes consecuencias en relación a la calidad: un drástico incremento del riesgo de introducir nuevos errores, y la dificultad para detectar dichas deficiencias.

Estas limitaciones implican que en toda investigación se debe plantear una premisa importante: durante el diseño del estudio es fundamental adoptar una serie de controles que garanticen la calidad del producto final y del proceso, lo cual supone necesariamente cuidar todos aquellos aspectos que inciden en el transcurso de la medición y en el correcto manejo de los datos. Sólo en algunos ámbitos de investigación, como en los ensayos clínicos, este proceso está sometido a una auditoría. En estos contextos, autores como Ronel, Varley y Webb (1993) proponen denominar dicho proceso como "gestión de datos" ("data management"). Lamentablemente, en otros muchos contextos, aunque conocen la necesidad de garantizar la calidad de la información que se maneja, éste aspecto aparece totalmente descuidado, hasta tal punto que a los datos se les "presupone" la calidad sin haber utilizado ninguna (o escasas) técnicas de control. Muy a nuestro pesar, gran parte de los estudios realizados en psicopatología adolecen de estos problemas.

En efecto, a pesar de los notables esfuerzos de algunos investigadores, en nuestra disciplina persisten graves deficiencias en el control de los datos debido a aspectos tales como (Freedland y Carney, 1992): (1) los obstáculos derivados del propio avance tecnológico (por ejemplo, los usuarios pueden no disponer del tiempo o recursos necesarios para adquirir y aprender el funcionamiento de los nuevos sistemas informáticos que van apareciendo); (2) la falta de incentivo para que los investigadores utilicen las nuevas técnicas de control de calidad de datos (especialmente cuando éstas suponen un trabajo tedioso, complicado o que les lleve mucho tiempo); y (3) la ausencia de planes de formación para profesionales orientados al estudio de los aspectos relacionados con la mejora en la calidad de la gestión de datos.

En otras palabras, el manejo correcto de los datos resulta en muchos casos excesivamente caro. Si bien en algunos proyectos se asegura la calidad de la información que se maneja, como los ensayos clínicos que disponen de los suficientes recursos y de la estricta legislación de la F.D.A., éste no es el caso de la mayor parte de los proyectos en Ciencias de la Salud. Sin embargo, por nuestra parte consideramos que aunque mantener un sistema efectivo de control de datos pueda resultar caro para un sector de la investigación con recursos limitados, o desconocedores de las herramientas informáticas más adecuadas, o con poca experiencia en su manejo, estas dificultades nunca deberían servir de excusa para dejar de incidir en aspectos fundamentales de control de datos. Todo investigador tiene el deber profesional de revisar rigurosamente sus datos y, cuando sea necesario, introducir todas aquellas técnicas de control que mejoren la gestión. Evidentemente, esta afirmación no pretende descargar toda la responsabilidad de la calidad del proceso de datos en cada investigador individual. Más bien al contrario, lo que proponemos es incentivar el uso de herramientas adecuadas y la realización de estudios sistemáticos que ayuden a clarificar la frecuencia, las causas y el impacto de los problemas de calidad en los procesos de medición y gestión. El objetivo no es otro que aproximar al mayor número de profesionales la idea que algunos autores preocupados por la calidad de los datos ya comenzaron a anunciar hace más de una década (Ember, 1986): "los resultados obtenidos en pequeños estudios con datos de

calidad pueden resultar mucho más interesantes y concluyentes que los obtenidos en grandes estudios con datos sin control de calidad o cuestionables en lo que se refiere a este aspecto".

Tradicionalmente, los procesos orientados a mejorar la calidad de los datos solían limitarse a la detección, revisión y corrección de los valores supuestamente erróneos ya grabados en la base de datos, y muchos investigadores consideraban que los aspectos más significativos relacionados con la calidad quedaban perfectamente resumidos en un número: la fiabilidad. Sin embargo, en la actualidad se asume que el estudio de la calidad conlleva mucho más que comprobar la adecuación de la información registrada, y que la fiabilidad, aun siendo un aspecto crucial de la información, no es el indicador más adecuado de la calidad cuando se presenta de forma única y/o aislada (Rowley, 1989). En este sentido, por ejemplo, los estudios más recientes emprendidos en esta línea conllevan procesos complejos para garantizar la precisión o exactitud de la información que se registra (Verter, 1990), la ausencia de valores omitidos (o desconocidos), la consistencia de los datos, la actualización (grado en que se ajustan al período estudiado), la unicidad y la validez (DISA, 1997).

Asimismo, en la actualidad está claramente establecido que para determinar la consistencia de los datos y los procedimientos adecuados de gestión se pueden adoptar diferentes puntos de mira. Por ejemplo, en 1989 Groves clasificó las potenciales fuentes de error que pueden atentar contra la calidad los datos de encuesta en tres categorías básicas (Groves, 1989): errores debidos a la no observación, errores debidos a una observación deficiente y errores de gestión o manejo de la información. En la primera categoría, el autor incluye, fundamentalmente, los errores de muestreo y los debidos a la dificultad para contactar y evaluar los sujetos inicialmente seleccionados para el estudio. En la segunda categoría incluye todos aquellos errores que pueden ser debidos al entrevistador, al entrevistado, al instrumento de evaluación y a la forma de obtener los datos. Finalmente, el autor clasifica en la categoría de los errores de gestión aquellas deficiencias debidas a la captura, la grabación y la manipulación de los datos para su posterior análisis estadístico. El texto de Groves (1989) aborda, básicamente, las fuentes de error debidas a la no observación y a la observación.

La clasificación de las fuentes de error que pueden incidir en los datos se puede abordar desde otras perspectivas bien distintas. Así, por ejemplo, Saris (1991) no contempla tan extensamente cuáles pueden ser las causas potenciales de error, sino que centra gran parte de su atención en valorar los recursos que ofrece el ordenador para eliminar dichas fuentes de error. En este sentido, podríamos decir que este autor abarca tanto los errores de no observación como los de medida y los de gestión, en un intento por decrementar tanto como sea posible su influencia.

Los aspectos relacionados con la calidad de los datos, enfocados desde la vertiente de la gestión informática, ocupan un espacio fundamental en los trabajos que nuestro equipo de investigación está llevando a cabo durante los últimos años. Por esta razón, la parte empírica de este proyecto se centra en la evaluación de los errores debidos a la observación y a la gestión de datos. Paralelamente, se presentan las estrategias que consideramos fundamentales para evitar, tanto como sea posible, dichas fuentes de error.

Este capítulo, de carácter teórico, se divide en dos apartados. En el primero se revisan las fuentes básicas de error recogidas en la literatura en relación al proceso de observación en psicopatología. En el segundo apartado se exponen los aspectos referidos a la calidad de la gestión de datos clínicos que consideramos fundamentales en el ámbito de las Ciencias de la Salud, adoptando como marco de referencia la teoría del diseño de bases de datos relacionales. El núcleo fundamental de esta parte del capítulo se basa en la reducción de errores de investigación a través de la adopción de estrategias de control y de las correspondientes restricciones.

## **3.2. FUENTES DE ERROR RELACIONADAS CON LA OBSERVACIÓN**

En los siguientes subapartados recogemos las fuentes básicas de error que se han relacionado con la calidad de la observación en los estudios de investigación que se llevan a cabo en Ciencias de la Salud, particularmente en psicopatología, así como también las estrategias fundamentales de control (Groves, 1989). Los siguientes subapartados recogen las principales variables de error que se han relacionado con el entrevistador, el entrevistado, la herramienta de evaluación y la forma en que se realiza dicha evaluación. Previamente, sin embargo, querríamos destacar que la determinación de dichos factores no resulta, en absoluto, una tarea sencilla. Las dificultades son todavía mayores cuando la entrevista es personal (participan un entrevistador y un entrevistado), ya que en estos casos lo que se produce es una interacción social en la que participan constantemente dos interlocutores. En efecto, la evaluación personal se caracteriza por un diálogo en el que tanto el entrevistador como el entrevistado juegan un papel ciertamente complejo (Saris, 1991). Así, por ejemplo, en el primer caso, el entrevistador debe obtener y asegurar la cooperación de los pacientes a lo largo de todo el proceso de evaluación, introducir aspectos que aseguren e incrementen la motivación de los sujetos, proporcionarles la información necesaria para obtener de ellos las respuestas deseadas (formular las preguntas, las opciones de respuesta, prestar ayuda cuando es requerida, dar instrucciones, etc.), crear un ambiente de confidencialidad en el que los informantes se expresen con la máxima libertad, seleccionar sólo la información relevante (descartando el resto de contenidos), y codificar las respuestas de acuerdo a los criterios establecidos en el instrumento de evaluación que utiliza. Por su parte, el informante debe interpretar las preguntas, elaborar las respuestas y formularlas (habitualmente en lenguaje verbal o escrito).

### **3.2.1. Fuentes de error debidas al observador/entrevistador**

Cuando la evaluación psicológica se realiza a través de la mediación de un entrevistador, todas las variables que se relacionan con dicho observador podrían constituir potenciales fuentes de error que atentan contra la calidad de los datos y de los resultados de la investigación. Asimismo, puesto que la evaluación psicopatológica que se realiza de forma personal suele atribuir gran importancia a la manera en que se conducen las entrevistas, la literatura ha considerado fundamental el estudio de las variables propias de los entrevistadores y de la interacción de dichos factores con el resto de elementos del proceso de evaluación.

Las fuentes de variabilidad debidas al entrevistador que pueden incidir en el proceso de evaluación y diagnóstico son numerosas. Edelbrock y Costello (1990), por ejemplo, destacan la importancia que tiene la forma de buscar y obtener la información de los entrevistados. En este sentido, una de las variables más estudiadas ha sido el grado de entrenamiento de los profesionales que conducen las entrevistas, ya que se ha postulado que la falta de experiencia en los protocolos de evaluación podría dificultar la capacidad para formular las preguntas de forma literal, en la secuencia correcta, y a un ritmo adecuado (ni demasiado despacio ni demasiado rápido). El entrenamiento permitiría que los resultados de la investigación para la que se utilizan unos determinados instrumentos de evaluación fueran fiables, y no la consecuencia de las interpretaciones y/o valoraciones particulares de cada evaluador respecto a dichos protocolos. Para lograr este objetivo se programan diferentes sesiones de entrenamiento, que suelen consistir en la observación y el registro de las respuestas facilitadas por casos reales y/o ficticios, utilizando cintas audiovisuales o entrevistas en vivo (esta valoración se efectúa de forma conjunta por un evaluador experto y otro novato). En cuanto a la relación entre el nivel de entrenamiento y la calidad final de los datos, hay pocos estudios que midan sistemáticamente este efecto, aunque se suele considerar que cuanto mayor experiencia poseen los evaluadores en el manejo de un protocolo, es menor la inferencia que

realizan respecto a la sintomatología y mayor la estabilidad de sus valoraciones (Gibbon et al, 1981; Von Cranach y Cooper, 1972; Young et al., 1987b). Asimismo, los resultados de diferentes estudios también sugieren que no basta con el entrenamiento inicial de los entrevistadores, sino que la supervisión y el "re-entrenamiento" deberían formar parte integral de todo el proceso de estudio (Ezpeleta, 1995; Fowler, 1993; Fowler y Mangione, 1990; Hodges, 1993). Finalmente, se ha comprobado que los errores de medida disminuyen cuando se instruye a los entrevistadores para que amplíen las instrucciones que dan a los sujetos evaluados en relación a lo que se pretende medir en cada pregunta, y también cuando se les instruye para que ofrezcan "feedback" a los entrevistados sobre los esfuerzos que están realizando al facilitar sus respuestas y sobre la calidad de las mismas (Cannell, Groves, Magilavy, Mathiewetz y Miller, 1987).

También se han evaluado qué características personales, demográficas, situacionales, actitudinales o cognitivas del evaluador podrían afectar el proceso de razonamiento clínico y, por consiguiente, la calidad del psicodiagnóstico. En este sentido, sólo cabe destacar que las inconsistencias de los resultados obtenidos en gran parte de las investigaciones realizadas no aportan la suficiente evidencia empírica como para determinar cuáles de estas variables contribuyen de forma decisiva sobre dichos aspectos de calidad (Moldafsky y Kwon, 1994).

De otro lado, también resulta de notable interés evaluar cómo interaccionan las variables del entrevistador con otras variables del proceso, fundamentalmente con las del entrevistado y con las características del instrumento de evaluación. En cuanto al primer punto, los elementos fundamentales que se han estudiado en relación a la variabilidad de los entrevistadores han sido el sexo, la edad y el nivel educativo de los evaluados (Groves y Magilavy, 1986). Aunque los resultados obtenidos no permiten concluir que el sexo de los entrevistados sea una variable que interactúe con la actuación de los entrevistadores, la edad y el nivel educativo sí parece que se deben tener en cuenta. En concreto, los resultados sugieren que los sujetos con menor nivel educativo, los niños pequeños y los sujetos de edad avanzada, constituirían la población más vulnerable a los efectos de la actuación de los evaluadores (Groves, 1989; Groves y Magilavy, 1986). Finalmente, en relación a la interacción entre las variables del entrevistador y las características del instrumento de evaluación, algunos estudios sugieren que cuando se formulan preguntas de contenido personal o actitudinal, que son más largas, más difíciles de procesar y responder, y que incorporan más términos ambiguos, las respuestas y su correspondiente registro están más sometidos a la variabilidad del evaluador para formularlas, dar las correspondientes aclaraciones y registrar las respuestas (Collins y Butcher, 1982).

### **3.2.2. Fuentes de error debidas al entrevistado**

No es fácil determinar qué variables del entrevistado pueden influir en los errores de medida que se supone son consecuencia de las respuestas dadas por los sujetos evaluados. Gran parte de los estudios realizados en esta línea se han centrado en evaluar los componentes que determinan el proceso de elaboración y formulación de las respuestas. En este caso, muchas investigaciones se han orientado hacia la psicología cognitiva y el paradigma del procesamiento de la información como referentes teóricos, ya que estos enfoques podrían explicar qué factores básicos determinan la adecuación de las respuestas dadas por los sujetos. De otro lado, la psicología social también ha contribuido a formular hipótesis sobre los errores de medida que podrían ser consecuencia de los entrevistados: se ha demostrado que es un referente útil para conocer el impacto del contexto social en la conducta humana, otorga evidencia experimental sobre por qué algunos sujetos utilizan determinados elementos para dar interpretación y significado a las preguntas, permite describir qué circunstancias afectan los procesos de memoria, y facilita explicaciones sobre fenómenos tales como el de la "deseabilidad social" (Groves, 1989).

Sea cual sea el marco de trabajo, uno de los factores clave relacionados con la posibilidad de que se produzcan errores de medida como consecuencia de los entrevistados consiste en el grado de comprensión de las preguntas. En este sentido, cuando las preguntas se formulan de la misma forma a todos los evaluados, se debe asegurar que las diferencias en las respuestas no sean consecuencia del grado de comprensión y/o posterior interpretación por parte de los sujetos (Fowler, 1993). Para descartar esta posibilidad, el estudio del grado de comprensión verbal se ha abordado no sólo desde la psicología cognitiva, sino también desde las teorías psicolingüísticas, en un intento por conocer qué factores determinan el nivel de comprensión de los entrevistados cuando las preguntas se presentan de forma oral (Clark, Schreuder y Buttrick, 1983). En cualquier caso, la mayoría de los trabajos realizados postulan que tanto el nivel educativo de los sujetos como la edad podrían ser dos de los determinantes clave del grado de comprensión, y que estos factores son especialmente relevantes cuando la evaluación se efectúa en población infantojuvenil.

Como ya hemos apuntado en el capítulo 1 al hablar del uso de las entrevistas diagnósticas estructuradas en población infantil y juvenil, parece que algunas características de los sujetos (especialmente la edad) sí serían importantes cuando se miden la fiabilidad y la validez de los datos recabados durante la evaluación (Young et al., 1987b; Zahner, 1991), aunque su influencia no sólo se manifiesta cuando interactúan con el grado de comprensión e interpretación. Por ejemplo, aunque algunos trabajos sugieren que la edad no afecta la fiabilidad de las respuestas de los informantes más jóvenes o el grado de acuerdo con la información de los padres (Herjanic et al., 1975; Silverman y Rabian, 1995; Verhulst, Althaus y Berden, 1987; Weissman et al., 1987), los resultados obtenidos en la mayoría de los trabajos sí indican que la edad constituye una variable importante a la que habrá que someter algunas condiciones del proceso (Achenbach et al., 1987; Boyle et al., 1993; Brent, Zelenak, Bukstein y Brown, 1990; Chambers, Puig-Antich y Hirsch, 1985; Canino et al., 1987; Costello, 1989; Edelbrock et al., 1986; Edelbrock, Costello, Dulcan, Kalas y Conover, 1985; Ezpeleta, de la Osa, Doménech, Navarro y Losilla, 1995a, 1995b; Herjanic y Reich, 1982; Hodges, 1993; Hodges, Gordon y Lennon, 1990; Kashani, Orvaschel, Burk y Reid, 1985; Kazdin, 1988; Klein, 1991; Kolko y Kazdin, 1993; Loeber et al., 1990; Loeber et al., 1989, 1991; Rapee et al., 1994; Reich y Earls, 1987; Reich et al., 1982; Rubio-Stipec et al., 1994; Schwab-Stone et al., 1994; Silverman, 1991; Silverman y Eisen, 1992). Asimismo, las conclusiones aparecidas en trabajos publicados recientemente sugieren que los niños, incluso los más pequeños, son capaces de dar respuestas fiables cuando son entrevistados directamente en los servicios de psicología y/o psiquiatría, aunque también es cierto que para ello se deben cuidar especialmente otros factores relacionados con la evaluación, sobre todo los contenidos de los enunciados y la forma en que se presentan. En esta línea, los resultados más pobres se obtendrían cuando los niños más pequeños tienen que responder preguntas complejas (por ejemplo, a nivel gramatical), que incluyen conceptos temporales o que requieren juicio (Breton et al., 1995; Fallon y Schwab-Stone, 1994; Schwab-Stone et al., 1994; Valla, Bergeron, Bérubé, Gaudet y Georges, 1994). Asimismo, en relación a la edad y de las relaciones que guarda con otros elementos del proceso de evaluación, otros estudios han medido el grado de relación entre la edad cronológica de los sujetos y el tipo de trastorno evaluado. En un estudio considerado ya clásico, Herjanic et al. (1975) hallaron un alto acuerdo entre padres e hijos de 6 a 13 años en lo relativo a preguntas objetivas, aunque este valor era considerablemente menor para las cuestiones referidas al estado mental. Por su parte, Silverman y Eisen (1992) observaron que los niños de 6 a 11 años informaban de forma más fiable que los padres sobre síntomas de ansiedad, en concreto en lo que se refiere a fobia simple y ansiedad excesiva. En general, los estudios encuentran prevalencias más altas de trastornos externalizantes cuando se entrevista a los padres, y prevalencias más elevadas de trastornos internalizantes cuando se entrevista a los hijos (Boyle et al. 1993; Chambers et al., 1985; Costello, 1989; Edelbrock et al., 1986; Herjanic y Reich, 1982; Loeber et al., 1990;



Loeber et al. 1989, 1991; Rapee, Barrett, Dadds y Evans, 1994; Reich y Earls, 1987; Rubio-Stipec et al., 1994; Silverman, 1991). Estos resultados sugieren que cuando se realizan estudios en población infantojuvenil es necesario obtener información de diversas fuentes si se quiere asegurar la calidad de los diagnósticos. Como indica Hodges (1993), padres e hijos no son informantes intercambiables, y tampoco se debe considerar que la información dada por los padres constituya el diagnóstico de referencia ("gold standard"). De hecho, cuando el clínico adopta esta actitud y sólo atiende a la información dada por los progenitores o por los otros cuidadores principales corre el riesgo de subestimar (incluso hasta la mitad) la prevalencia de la psicopatología infantojuvenil (Bird, Gould y Staghezza-Jaramillo, 1992; Carlson et al., 1987; Costello, 1989; Klein, 1991; Leckman, Sholomskas, Thompson, Belanger y Weissman, 1982; Pedreira y Sánchez-Gimeno, 1992; Piacentini, Cohen y Cohen, 1992; Spitzer, 1983). En esta población, la falta de acuerdo entre las fuentes no debería entenderse como un factor que vulnera la calidad de los datos, ni tampoco como una limitación del protocolo de evaluación, sino como resultado de contemplar un mismo fenómeno sintomático desde diferentes perspectivas y en diferentes contextos (Achenbach et al., 1987; Klein, 1991; Loeber et al., 1990; McConaughy et al., 1988). En este sentido, la discrepancia constituiría un factor importante que aporta información relevante de cara al diagnóstico y al tratamiento.

Otro de los factores fundamentales relacionados con los errores de medida procedentes de los entrevistados consiste en el grado de recuerdo de los contenidos por los que se pregunta durante la evaluación. En este caso, se ha postulado que además del importante rol que juegan el nivel de atención y motivación de los entrevistados, existen dos variables fundamentales relacionadas con la calidad de las respuestas: el tiempo transcurrido desde que sucedió el evento por el que se pregunta, y el grado de importancia que el evaluado otorga a dichos sucesos. En el primer caso, los estudios realizados constatan que el nivel de recuerdo es más deficiente cuando las preguntas se refieren a hechos acontecidos en el pasado, y que las dificultades se incrementan cuanto más alejados en el tiempo están dichos sucesos (Duncan y Hill, 1985; Machiels-Bongaerts, Smitdt y Boshuizen, 1993). En el segundo caso, estudios empíricos constatan que los sujetos informan de forma más consistente y precisa sobre sucesos a los que otorgan mayor grado de significación o importancia (Brown, Rips y Schevell, 1985).

Finalmente, otro factor básico relacionado con los errores de medida del entrevistado es el proceso de toma de decisiones que realiza para elegir la respuesta mas apropiada frente a una determinada pregunta. En este punto, también ha recabado cierto interés el estudio del contenido y el estilo que utilizan los sujetos al dar sus respuestas. En cualquier caso, al hablar de las fuentes de variabilidad relacionadas con el entrevistado, la calidad de las respuestas se ha relacionado con dos elementos básicos: los efectos de "deseabilidad" y las tendencias de respuesta. En el primer caso, se ha desarrollado un ingente número de investigaciones que destacan cómo los sujetos tienden a dar respuestas "socialmente deseables" (Silver, Anderson y Abramson, 1986). En el segundo capítulo abordamos esta cuestión al hablar de la equivalencia experiencial y de la actitud de los evaluados (cuando comparamos las pruebas tradicionales y las nuevas versiones informáticas), y comprobamos que esta potencial fuente de error se reducía notablemente cuando el sujeto realizaba la entrevista directamente a través de un ordenador (Alhberg et al, 1996; Duffy y Waterton, 1984; Erdman et al., 1987; Ferriter, 1993; Hinkle et al., 1991; Kovak et al., 1994, 1996; Levine et al., 1989; Meszaros et al., 1995; Petrie y Abell, 1994; Robinson y West, 1992). En el caso de las tendencias de respuesta, algunos trabajos apuntan que los sujetos tienden a dar respuestas en la misma dirección cuando responden las preguntas de los cuestionarios (Schuman, Presser y Ludwig, 1981), lo cual ha hecho que muchas de las pruebas de evaluación tradicionales incorporen ítems formulados de forma directa e inversa. El fenómeno de las tendencias de respuesta también se observa en sujetos que durante las entrevistas aprenden que dando una determinada respuesta

se reduce el tiempo requerido para la evaluación. Este sería el caso de los pacientes que sistemáticamente contestan "no" a las preguntas de salto de una entrevista estructurada, independientemente de su estado psicopatológico real, ya que de este modo no se les formulan las preguntas contenidas en dichos saltos y la entrevista acaba antes. Este fenómeno, que en ocasiones se denomina de "atenuación", también se observa cuando las preguntas que se le formulan al entrevistado son poco claras, cuando rebasan su capacidad de comprensión, cuando requieren de interpretaciones complejas o simplemente cuando los evaluados consideran que sólo deben dar información nueva que nunca antes hayan aportado (Ezpeleta, 1995; Robins, 1985; Hodges y Zeman, 1993). Finalmente, aunque en la práctica es menos habitual, también es posible que los sujetos tiendan a dar sistemáticamente respuestas afirmativas a las preguntas que se les formulan sobre su estado psicopatológico, por ejemplo en un intento de llamar la atención al entrevistador.

### **3.2.3. Fuentes de error debidas al instrumento de evaluación**

Las características del protocolo utilizado para la evaluación juegan un papel fundamental en la calidad de las respuestas obtenidas por parte de los sujetos y en su correspondiente registro. Aunque existen pocos estudios que aborden con precisión qué efectos puede tener cada elemento del formulario sobre el total de errores de medida de una investigación, lo que sí ha quedado claramente establecido a lo largo de los últimos años es la necesidad de estandarizar. En este sentido, la introducción de instrumentos de evaluación con contenidos estructurados y con formatos previamente definidos para formular y registrar respuestas, se ha relacionado con un incremento en la fiabilidad de los diagnósticos que se definen en psicopatología (Ezpeleta, 1996). Se ha hipotetizado que este hecho puede ser debido a que la estandarización conlleva una notable reducción en la variabilidad del intercambio pregunta-respuesta. De este modo, cuanto mayor sea el grado de estandarización del protocolo, mejor definidos estén los ítems y menor sea el nivel de inferencia requerido de los evaluadores para conducir la entrevista, mayor fiabilidad merecerá la información diagnóstica que se maneja (Ezpeleta, 1990; Herjanic, 1984).

También se ha hipotetizado que las fuentes principales de error debidas al tipo de cuestionario se relacionan con los reactivos utilizados, y se concretan en el tipo de lenguaje, la estructura de las preguntas y la forma requerida para que los sujetos faciliten las respuestas (Groves, 1989).

En cuanto a los efectos del tipo de lenguaje utilizado en cada pregunta, además del papel fundamental que juegan los términos utilizados de cara al grado de comprensión de los evaluados, se ha hipotetizado que las fuentes de error se incrementan con el uso de términos vagos o demasiado comunes (Groves, 1989), ya que en estos casos los enunciados son objeto de interpretación libre por parte de los sujetos (Fowler, 1992; Oksenberg, Cannell y Kalton, 1991). Sin embargo, todavía no se dispone de la suficiente evidencia empírica como para poder predecir qué expresiones o términos se relacionan con un incremento del número de errores; lo único claro en este sentido es que pequeños cambios en la redacción de las preguntas suelen conllevar diferencias notables en la calidad de las respuestas.

De otro lado, en relación a los efectos debidos a la estructura de las preguntas, parece claro que la longitud de los enunciados, el tipo de respuesta requerida -abierta "versus" cerrada-, el número y orden de presentación de las alternativas de respuesta, el tipo de contenido -objetivo "versus" subjetivo, o externalizante "versus" internalizante-, y el orden de presentación de las preguntas, son los elementos más estrechamente relacionados con la calidad de las respuestas y de su correspondiente registro (Collins y Courtenay, 1985; Fowler, 1993; Groves, 1989; Miller, 1984; Saris, 1991). En este sentido, los resultados más deficientes en cuanto a calidad se obtendrían con preguntas muy largas, de opción de respuesta abierta, y cuyos contenidos sean evaluados como intrusivos por parte del sujeto evaluado. En los dos

primeros casos el entrevistado puede encontrar dificultades para determinar cuál es el objetivo de la pregunta y cuáles constituyen las respuestas más adecuadas en cada caso. En el segundo, si el sujeto considera que la pregunta es intrusiva, puede dudar sobre el grado de confidencialidad de la entrevista y las repercusiones de facilitar una respuesta, o incluso puede manifestar un cierto grado de pudor por mostrar contenidos que socialmente se consideran personales y/o indeseables.

Finalmente, algunos trabajos también se han interesado por los efectos que tiene la forma en que se requiere que los sujetos faciliten la información sobre la calidad de los datos. En este sentido, se ha sugerido que el grado de comprensibilidad y la claridad de presentación de las preguntas son factores más importantes que la forma en que se pide a los sujetos que respondan (French y Beaumont, 1987), aunque algunos estudios más recientes apuntan que dicha hipótesis carece de la suficiente evidencia empírica.

Antes de finalizar este apartado queríamos destacar algunos trabajos que indican que cuando la evaluación implica niños pequeños, es básico considerar los aspectos relacionados con el tipo de lenguaje, la estructura de las preguntas o la forma de respuesta (Breton et al., 1995; Fallon y Schwab-Stone, 1994; Schwab-Stone et al., 1994; Valla et al., 1994). Aunque todavía no disponemos de datos definitivos sobre el efecto específico de estos factores sobre la calidad de las respuestas, se ha postulado que los elementos más importantes a tener en cuenta son: la complejidad gramatical, la presencia de constructos abstractos en los enunciados (como afectos o tiempo) y la necesidad de que se emitan juicios sobre comportamientos.

#### **3.2.4. Fuentes de error debidas a la forma de recogida de datos**

Tradicionalmente, la evaluación psicopatológica se ha realizado de forma personal, es decir, a partir de un entrevistador encargado de conducir la entrevista con un entrevistado. Este tipo de evaluación constituye un acto de comunicación entre dos personas, con las ventajas e inconvenientes que suponen dichas interacciones. En el caso concreto del proceso de evaluación, lo que sí es cierto es que la comunicación se desarrolla de forma muy estructurada, debido a los requisitos que necesariamente impone dicho proceso. Esto supone que algunas normas de comportamiento propias de la comunicación entre dos personas (por ejemplo, las reacciones frente a los comentarios del interlocutor), no se manifiestan ni del mismo modo ni en la misma intensidad.

En las últimas décadas, se han propuesto otras formas diferentes de evaluación. Entre las más destacadas se encuentran la evaluación a través del teléfono y la autoadministración de pruebas a través del ordenador. Las diferencias fundamentales entre las entrevistas personales y las realizadas a través del teléfono o del ordenador consiste en que éstas últimas utilizan sólo una parte de los canales de comunicación habituales en las interacciones sociales (Sykes y Collins, 1987). En este sentido, se ha postulado que las dos nuevas modalidades podrían tener serias limitaciones de cara al establecimiento adecuado de los diagnósticos psicológicos, puesto que limitarse al registro de un tipo de información (en el caso del teléfono, por ejemplo, la que llega al entrevistador vía auditiva) hace que el entrevistado adopte normas de conducta y actitudes diferentes, y que el profesional en psicología no disponga de otros elementos que tradicionalmente se han considerado casi fundamentales de cada al correcto diagnóstico de los trastornos. Los defensores de esta postura consideran, además, que el canal visual es importante para valorar y verificar aspectos tales como el estado afectivo del interlocutor (por ejemplo, interpretando las expresiones faciales o los gestos) o el estatus económico y social (por ejemplo, en función del estilo con el que los sujetos visten). De otro lado, algunos autores también han argumentado que la falta de información procedente del canal visual no es tan determinante, y que cuando los sujetos evaluados proporcionan respuestas directamente al ordenador se pueden medir

simultáneamente otros aspectos psicofisiológicos (como la frecuencia cardíaca, la latencia de respuesta o la presión que se ejerce al teclear las respuestas) que podrían servir de ayuda tanto para la formulación de hipótesis diagnósticas como para modular el ritmo o la forma de administrar los reactivos (Erdman et al., 1985; Slack, 1971; Temple y Geisinger, 1990). Lamentablemente, esta última línea de investigación no ha gozado de la atención que nosotros consideramos que se merece. La mayoría de los programas diseñados con fines diagnósticos en psicología no suelen incluir la valoración de contenidos no verbales obtenidos directamente del sujeto evaluado, y hasta la fecha no tenemos constancia de ningún programa que sea capaz de captar todas aquellas respuestas no verbales que los entrevistadores humanos captan y utilizan durante el proceso de evaluación. También es cierto que tampoco sabemos en qué grado esta información puede o no ser de utilidad de cara al diagnóstico. Algunos autores sugieren que tal vez estos datos sólo son útiles (o más fiables, o más válidos) para la evaluación de algunos trastornos particulares, como por ejemplo los de tipo psicótico (Lewis et al., 1988). Otros autores, sin embargo, sugieren que estas medidas podrían servir para mejorar la comprensión de los procesos y componentes que subyacen los constructos psicopatológicos objeto de evaluación, y también para obtener información por vía indirecta sobre aspectos tales como la fatiga, el grado de motivación, la habilidad cognitiva, los patrones de respuesta de los sujetos e, incluso, las deficiencias asociadas a posibles trastornos de la lectura (Golu y Bogatu, 1993; McLennan et al., 1988).

En cualquier caso, a pesar de las posibles limitaciones que podría conllevar el uso de las nuevas tecnologías, un gran número de trabajos se han esforzado en señalar que por encima de tales deficiencias existen numerosas ventajas en lo que se refiere a registro y gestión de datos. Así, por ejemplo, se ha postulado que la evaluación con ordenador resulta más económica y conlleva menos tiempo que la evaluación personal con un entrevistador. Esto permitiría que gran parte de los profesionales dispusieran de más tiempo para concentrarse en otros aspectos relacionados con la calidad de la investigación que llevan a término (Bloom, 1992; Groves, 1989; Moreland, 1990b; Sykes y Collins, 1987). Asimismo, y aunque parezca evidente, el uso del ordenador como herramienta básica y única para la recogida de los datos implica la casi completa desaparición de los errores debidos al entrevistador. En este sentido, lo que todavía resulta indispensable comprobar es que con la desaparición de la figura del entrevistador no desaparezcan también los posibles efectos positivos que su presencia aportaba de cara a la calidad del estudio.

De otro lado, en el capítulo 2 también indicamos que, en general, la actitud de los sujetos evaluados es más favorable cuando se les solicita que faciliten respuestas directamente al ordenador que con entrevistas personales. En este sentido, por ejemplo, los resultados obtenidos en múltiples estudios sugieren que los evaluados muestran un alto nivel de satisfacción cuando responden cuestionarios autoadministrados con ordenador (Baker, 1992; Burda et al. 1991; Dignon, 1997; Farrell et al., 1987; Harrell et al., 1987; Honaker et al., 1988; Kight-Law et al., 1989; Krieger y Dlugosch, 1991; Lucas et al., 1977; Plutchik y Karasu, 1991; Rosenman et al., 1997; Rozensky et al., 1986; Sawyer et al., 1990, 1991), que facilitan más información y más precisa sobre aspectos que consideran personales o íntimos (Johnston y Walton, 1995; Kobak et al., 1996; Rolnick et al., 1989; Sawyer et al., 1991), y que perciben más rápido el proceso de evaluación (Gottschalk et al., 1997; Honaker, 1988; Kobak et al., 1996). Estos resultados, en conjunto, aportan indicios sobre las mejoras que se pueden conseguir en relación a la calidad de las respuestas, y sugieren que los nuevos modos de evaluación incrementan la fiabilidad de los datos y decrementan el volumen de información faltante (valores desconocidos).

También se ha postulado que existen diferentes factores que podrían interactuar con la forma en que se recoge la información, y que por lo tanto no deberían ser obviados cuando se estudia la relación entre la forma en que se lleva a cabo la evaluación y la calidad de la gestión de datos: las características biográficas de los evaluados (especialmente el sexo, la edad y el

nivel educativo), el tipo y el grado de alteración psicológica, la forma requerida para introducir las respuestas (teclado, pantallas táctiles, etc.), la forma de presentación de las preguntas, el grado de experiencia previa con ordenadores, o incluso la habilidad informática. Entre las variables más estudiadas de la lista anterior se encuentran los factores biográficos y psicopatológicos de los sujetos, y parece que ni el sexo ni el tipo o el grado de psicopatología guardan relación con la forma de administrar las pruebas. En cambio, la edad y el nivel cultural sí parecen jugar un papel importante, y los resultados de distintos estudios de investigación sugieren que las personas más mayores (Burke et al., 1987; Spinhoven et al., 1993) o con menor nivel educativo (Skinner y Pakula, 1986) podrían mostrar más efectos debidos al modo en que se recogen los datos. Estos resultados no sólo serían consecuencia del grado de experiencia o habilidad con las nuevas tecnologías, sino también de las cogniciones y expectativas que mantiene esta población en relación al proceso de evaluación, o incluso a los déficit sensitivos habituales en personas de mayor edad. En cualquier caso, también es importante destacar que los resultados hallados en esta línea de trabajo no siempre son concluyentes. Así, por ejemplo, Cannet et al. (1987), hallaron una serie de diferencias difíciles de interpretar como consecuencia de la variable edad, lo cual sugería que dicho estudio adolecía de problemas metodológicos o que tales diferencias simplemente no existían.

Finalmente, también se ha hipotetizado que parte de las mejoras en la calidad observadas con el uso de los nuevos procedimientos de evaluación, especialmente cuando se usa la tecnología informática, podrían ser consecuencia de los esfuerzos que los investigadores emprenden en el desarrollo de los nuevos protocolos (Godoy, 1996; Groves, 1989). Así, por ejemplo, es habitual que cuando se usa el ordenador para obtener directamente las respuestas por parte de los sujetos, se utilicen preguntas más sencillas con conceptos más fáciles de comprender, que las entrevistas sean más cortas, o que se elaboren diseños que favorezcan el mantenimiento de aspectos atencionales, motivacionales y/o perceptuales. En relación con esta línea de investigación nos encontramos los trabajos de la "Human Computer Interaction" (HCI), y los que proceden de la construcción de reactivos para entrevistas y cuestionarios (Carroll, 1993; Dillon, 1983; González, 1993; Johnson, 1992; Lalomia y Sidowski, 1993; Pocius, 1991; Schneiderman, 1992; Wallace y Anderson, 1993). En ambos casos, los profesionales procuran obtener evidencia empírica sobre qué elementos de los cuestionarios y los modos de interacción aseguran la atención y la motivación del sujeto, a la vez que inciden en aspectos perceptuales favorecedores del procesamiento de la información y la elaboración de las respuestas.

### **3.3. FUENTES DE ERROR RELACIONADAS CON LA GESTIÓN DE LOS DATOS**

En lo que resta del capítulo revisamos qué fuentes de error básicas se relacionan con la calidad de la gestión de datos, y cuáles han sido las estrategias básicas que se han emprendido para mejorar su calidad. Para desarrollar estos aspectos se adopta la teoría del diseño de bases de datos relacionales como marco de trabajo. Previamente, en el primer subapartado plantharemos los elementos básicos que se deben tener presente cuando se aborda el estudio de las bases de datos.

### **3.3.1. Base de datos, sistemas de bases de datos y sistemas de gestión de bases de datos**

#### **Base de datos**

Una base de datos designa un conjunto de información organizada de acuerdo a algún criterio. La definición de cualquier base de datos tiene como objetivo proveer de la estructura y especificaciones que se requieren para que la entrada de datos se efectúe de forma precisa, eficiente y consistente, a la vez que facilita la posibilidad de realizar consultas<sup>4</sup>, transformaciones, relaciones y listados de la información de forma flexible a los requisitos de los usuarios. En particular, cuando se alude a base de datos en términos informáticos, se está designando un conjunto de información que está contenida y organizada de acuerdo a algún criterio en los sistemas de memoria del ordenador, y que es susceptible de ser consultada y/o manejada por uno o varios usuarios.

Cuando se realiza una investigación se pueden definir una o varias bases de datos. La literatura y la práctica ofrecen diferentes razones para usar una única base en la que los datos se almacenan sólo una vez (independientemente de que el estudio se efectúe en diferentes centros o espacios), en lugar de mantener varios ficheros separados, posiblemente no disjuntos (Date, 1989; Frost, 1989). Así, por ejemplo, disponer de una sola base ahorra recursos en el sistema (siempre claro está que los datos se reúnan y almacenen sin duplicidades), permite que los datos se aprovechen mejor cuando se requiere que diferentes usuarios y/o aplicaciones deban utilizarlos, y permite que se cometan menos inconsistencias (en aspectos tales como la actualización de las copias de seguridad o el chequeo para detectar que no hay incompatibilidades en la información).

#### **Sistemas de bases de datos**

Un sistema de bases de datos es un conjunto de recursos cuya responsabilidad colectiva incluye: (1) almacenar los datos; (2) mantener la seguridad de la base de datos reforzando las medidas de privacidad e integridad y también proporcionando la reserva necesaria para prever fallos en equipos/programas; y (3) proporcionar a los usuarios usos con las rutinas necesarias de entrada/salida, de tal forma que se pueda acceder fácilmente a la base de datos cuando el estudio lo quiera.

El punto de partida para el diseño de un sistema de bases de datos consiste en una descripción abstracta y general de esa parte del universo que los datos representan. Dicha descripción abstracta se denomina "esquema conceptual", y contiene una lista de los tipos de entidades implicadas, una lista de las correspondientes relaciones y una lista de las restricciones de integridad.

La información de la base de datos se organiza en archivos a modo de tablas bidimensionales, donde las filas (también denominadas "registros" desde el punto de vista del archivo de datos, o "casos" desde el punto de vista de la estadística) constituyen cada una de las entradas de la base de datos, y las columnas (también denominadas "campos" desde el punto de vista del archivo, o "variables" desde el punto de vista de la estadística) representan cada una de las características de la entidad que se almacena (Losilla y Navarro, 1997).

En este trabajo damos por supuesto que un componente principal del sistema de bases de datos es un ordenador, y que la base de datos reside en dicho sistema.

---

<sup>4</sup> En este trabajo entendemos por consulta el medio característico utilizado por los usuarios de los sistemas gestores de bases de datos a la hora de formular y responder preguntas sobre los datos contenidos en dichas bases. En este sentido, las consultas serían el procedimiento estándar para seleccionar y organizar exactamente la información que se desea visualizar, tanto a nivel de campos como a nivel de registros.

## Sistemas de gestión de bases de datos

Un sistema de gestión de bases de datos (SGBD) es una colección de procedimientos, ayudas de documentación, lenguajes y programas para facilitar el diseño e instalación de sistemas de bases de datos. Generalmente, un SGBD está basado en una estructura de almacenamiento de propósito general que puede ser adaptada para cumplir los requerimientos de varios sistemas de bases de datos. Normalmente, el sistema viene equipado de subsistemas de privacidad, integridad y recuperación, que también pueden adaptarse para requerimientos particulares (Gillman y Gillian, 1991; Litton, 1991; Jackson, 1990).

En términos informáticos, el SGBD designa el software o sistema informático concreto que cumple con el objetivo básico de administrar la base de datos. Entre las operaciones básicas que se le imputan destacan: la organización y estructuración de los datos, la captura y mantenimiento de la información, y el manejo de las solicitudes de acceso (Losilla y Navarro, 1997). Desde el punto de vista del usuario, estas operaciones se concretan en la adición, modificación y eliminación de la información contenida en la base, la localización y recuperación de datos que cumplen una determinada condición, la ordenación de la información según uno o más campos, y la importación/exportación de datos.

Lamentablemente, a pesar de los enormes progresos en software y hardware de los últimos años, todavía existen relativamente pocos sistemas que permitan definir y manejar de forma sencilla bases de datos con estructuras complejas. En este sentido, la mayoría de los programas existentes en el mercado adolecen de serias limitaciones, entre las que destacan (Frost, 1989): (1) el requisito de una extensa programación cuando se desean representar relaciones complejas; (2) la dificultad para el manejo de información textual; (3) la limitación del poder expresivo de los diversos lenguajes de especificación proporcionado por el SGBD para representar la estructura de los datos; (4) la tarea del diseñador de un SGBD se complica por la necesidad de estar familiarizado con la terminología de la estructura de los datos que utiliza el SGBD; (5) los usuarios finales del sistema deben estar plenamente familiarizados con la estructura de los datos que subyace bajo el SGBD; (6) la independencia de los datos no está garantizada; (7) la verificación de la integridad (adaptación de la base de datos con las restricciones derivadas de lo que se conoce como permitido en aquella parte del universo que está representada por los datos) no es automática; (8) la recuperación deductiva no está garantizada; (9) gran parte de los SGBD se basa en equipos inadecuados; y (10) las interfaces del usuario final no acostumbra a ser sencillas. A pesar de que este listado de dificultades fue enunciado hace caso 10 años, nos gustaría señalar un aspecto que, a nuestro juicio, resulta todavía fundamental: las ventajas que se consiguen a nivel de manejo y calidad de gestión a través de diseños de bases de datos podrían verse contrarrestadas si el gestor que se escoge no resulta fácil de utilizar, tanto en sus prestaciones para definir la base de datos y las pantallas de entrada de datos, como en el tiempo requerido para el entrenamiento de los futuros usuarios. Se debe tener presente que los usos más extendidos de los SGBD se relacionan con aspectos fundamentales de la investigación, como la definición de la estructura de los datos del estudio, la entrada de la información y su posterior manejo. Por lo tanto, para que la estructura definida opere de forma eficiente y sencilla, es importante que el sistema facilite al usuario, tanto como sea posible, todas estas tareas. En cualquier caso, conviene señalar que en los últimos años el reconocimiento de las limitaciones de los enfoques convencionales para el trabajo con bases de datos ha dado lugar a una reconsideración de la arquitectura de los SGBD. Los programas resultantes son, de alguna manera, más sencillos que los convencionales.

Los SGBD se basan en modelos o estructuras de datos de propósito general que se pueden adecuar para cumplir los requisitos de cada aplicación particular. En la actualidad se reconocen diferentes estructuras de datos para almacenar la información, entre las que destacan: el modelo jerárquico, el modelo de red, el modelo relacional, el modelo relacional

binario, el modelo semántico y el modelo orientado a objetos. Si atendemos al número de aplicaciones disponibles en el mercado y en la comunidad científica, se observa que los enfoques jerárquicos y de red están en claro declive. En cambio, el enfoque relacional es actualmente el más extendido, tal vez como consecuencia de que este enfoque gana a los anteriores en potencia y en número de prestaciones.

A la hora de elegir un SGBD es importante que el usuario conozca el modelo de datos subyacente. Con objeto de comentar algunas ventajas e inconvenientes entre bases de datos relacionales y no relacionales, vamos a presentar las características fundamentales de los modelos de gestión de datos más utilizados en los últimos años (Frost, 1989; Losilla y Navarro, 1997):

- **Modelo de datos jerárquico:** considera que el universo consta de entidades, atributos y relaciones. Las entidades están relacionadas a otras entidades en estructuras arborescentes.
- **Modelo de datos de red:** hacen la misma distinción entre entidades y atributos que en la perspectiva jerárquica, pero las estructuras permitidas que vinculan entidades son mucho menos restrictivas. De este modo, las entidades pueden estar relacionadas con cualquier número de otras entidades.
- **Modelo relacional:** se basa en el concepto de relación de la teoría matemática de conjuntos, lo cual hace que sea posible aplicar todas las formulaciones de dicha teoría al sistema de gestión de los datos. La disponibilidad de diferentes fuentes de información y de relaciones entre dichas fuentes constituye un ejemplo clásico para el uso de sistemas de gestión de bases de datos relacionales. El concepto central de esta tecnología radica en la descomposición de cada aplicación en series de matrices simples o rectangulares (una para cada entidad), cada una de las cuales se relaciona con las otras. Asimismo, el modelo relacional se estructura en diferentes niveles de representación: un nivel externo contenedor de las vistas o consultas, un nivel conceptual o lógico contenedor del archivo-catálogo con la información sobre todos los elementos de la base y de sus relaciones, y un nivel interno encargado de almacenar físicamente los archivos.
- **Modelos semánticos:** se han propuesto como alternativas al modelado clásico con el objeto explícito de aportar constructos que permitan al diseñador capturar más componentes semánticos de un problema particular. La premisa básica de este tipo de representación supone que una base de datos basada en estructuras naturales es más simple de diseñar y modificar que una base de datos que requiera de la traslación de los objetos de medida en constructos especificados artificialmente. En cualquier caso, los argumentos utilizados en favor de los modelos de datos semánticos no gozan de apoyo universal.
- **Modelos de datos orientados a objetos:** constituyen uno de los desarrollos más recientes y prometedores dentro del modelado semántico. Esta aproximación permite a los usuarios representar objetos que poseen una correspondencia uno-a-uno con las entidades de un dominio específico. Entre las principales ventajas de este paradigma destaca su flexibilidad.

En la actualidad existen diferentes trabajos que exponen las ventajas de utilizar el sistema relacional de bases de datos (Losilla y Navarro, 1997; Jackson, 1990). Así, por ejemplo, se ha considerado que dicho modelo ahorra espacio en disco (fundamentalmente por la existencia de tablas relacionadas que evitan la introducción de información repetida), que disminuye el tiempo de registro de la información, que evita errores de escritura y/o transcripción de datos (por ejemplo, se impide que una misma información se introduzca con



diferentes variantes), que incorporan procedimientos sistemáticos que facilitan el diseño de la base de datos, y que disponen de sofisticados lenguajes que permiten a los usuarios realizar complejas búsquedas de información con un mínimo esfuerzo (tales como el "Structured Query Language" SQL o el "Query By Example" QBE). Este clima ha favorecido que hoy en día la mayoría de los SGBD diseñados para microordenadores adopten la estructura relacional, ya que de este modo se considera que se facilita la interacción usuario-sistema.

Asimismo, y a pesar de la popularidad alcanzada por el modelo relacional, este marco de trabajo también ha sido objeto de algunas críticas (Batra y Srinivasan, 1992). Así, por ejemplo, se ha destacado el hecho de que la teoría relacional no ofrece ninguna indicación sobre cómo deben ser representados los objetos de estudio en términos de relaciones. También se ha postulado que los conceptos de comprensibilidad, eficiencia y consistencia del modelo no han sido demostrados con suficiente evidencia empírica.

Antes de finalizar este apartado también nos parece importante aportar una breve nota histórica. En este sentido, destacar que desde la década de los 70, numerosas organizaciones han concentrado gran parte de sus esfuerzos en el desarrollo de sistemas informáticos que les permitieran recoger datos utilizando protocolos estructurados (habitualmente cuestionarios autoadministrados o entrevistas personales o telefónicas). Aunque estos sistemas todavía se utilizan, en los últimos años empresas y organizaciones se están interesando por el desarrollo de sistemas que manejen grandes y complejas estructuras de datos. Por ejemplo, sistemas que permitan utilizar cuestionarios con relaciones más complejas que las simples jerarquías, usar más de un instrumento de evaluación en cada sujeto, manejar o asignar fuentes de datos a múltiples estudios simultáneamente, y crear y mantener grandes volúmenes de datos. En la actualidad, los sistemas de gestión de datos aparecidos en el mercado están alcanzando tal estado de sofisticación que gran número de investigadores trabajan en la posibilidad de integración de las diferentes tareas del procesamiento de la información en un único sistema. Éste debería encargarse de integrar las fases de preparación de la documentación, del diseño y administración de los instrumentos de evaluación, del manejo y combinación de la información registrada y de su posterior análisis estadístico. En este sentido, los límites entre los sistemas diseñados para la captura de los datos, su manejo y el posterior análisis se están comenzando a difuminar, en la medida en que muchos paquetes centrados en una de estas tareas adquieren también capacidades en otras áreas sin necesidad de sacrificar la integridad de sus aplicaciones originales. Tal vez el ejemplo más claro lo constituyen los programas de análisis estadístico que implementan opciones para entrar los datos y manejar estructuras no rectangulares; y también las bases de datos, que suelen realizar cálculos o resúmenes estadísticos de la información que contienen. En relación a esta nueva tecnología, también se debe indicar que si bien es cierto que esta tendencia expansiva parece afianzarse en los últimos años, no parece probable que en breve dispongamos de un único sistema capaz de integrar las tres actividades básicas de la gestión de datos: recogida, manejo y análisis de la información. Tal vez la causa radique en que cada una de las fases del proceso de datos se caracteriza por unos requisitos propios y tiene también unas dificultades específicas, que a su vez suelen resolverse o abordarse ignorando simplemente cuestiones que son esenciales en otras fases. Evidentemente, considerar la posibilidad integradora es interesante en la medida en que plantea la necesidad de reformular el proceso de investigación en sí mismo. Efectivamente, la disponibilidad de sistemas integradores permitiría que muchos profesionales se dedicaran al cuidado de las cuestiones básicas relacionadas con la calidad de gestión de datos, ya que los otros aspectos complicados del proceso de gestión quedarían automatizados.

El uso de bases de datos relacionales participa, aunque sea parcialmente, de este último propósito. La disponibilidad de tales sistemas tiene como objetivo proporcionar los medios necesarios para que los investigadores mejoren aspectos de calidad del proceso de datos. La tarea, sin embargo no es sencilla. La definición y creación de una base de datos relacional requiere el diseño de una estructura lógica, en la que se describa tanto el fenómeno objeto de

estudio como sus relaciones (Todd, Jacobus y Boland, 1992). Esta estructura viene determinada por los requisitos de cada usuario (o si se prefiere, del estudio), los cuales dependen a su vez de la naturaleza del contexto donde la base de datos debe ser implementada y de los objetivos particulares de cada proyecto.

### **3.3.2. Cómo mejorar la calidad de la gestión de datos desde la óptica del diseño de bases de datos**

Una forma de enfocar la calidad de la gestión de datos consiste en tomar como modelo de trabajo el enfoque del diseño de bases de datos. Una de las razones fundamentales por las que hemos escogido esta aproximación como marco de trabajo consiste en la posibilidad de ubicar el diseño de la base de datos como una fase a la que se atribuye, al menos, tanta relevancia como al resto de fases que componen la evaluación clínica y el proceso de investigación en conjunto. Esto implica la necesidad de llevar a cabo un conjunto de operaciones destinadas a garantizar la sistematización de la captura de datos, la adecuada combinación de la información disponible y la correcta preparación de la matriz de datos para el posterior análisis estadístico. En particular, en el capítulo 5 de este proyecto presentamos el proceso seguido para informatizar una entrevista diagnóstica estructurada.

En este trabajo conceptualizamos el proceso de datos como una actividad compleja cuyas principales etapas de actuación son: el diseño de la base de datos (definición de la información a registrar, creación de la estructura de las tablas y especificación de las protecciones de registro), la grabación (captura) de la información, la preparación de la matriz de datos para el análisis estadístico (generación de variables, selección de casos y creación de tablas rectangulares) y el análisis de datos. En nuestra opinión, esta propuesta es compatible con la que realiza Litton (1991), quien considera que desde una perspectiva netamente funcional el diseño de la base de datos constituye un proceso (iterativo) basado en criterios normativos que establecen tanto la estructura de la base de datos (características de la información objeto de estudio y relaciones) como las restricciones, las consultas permitidas, el diseño de la interfaz sistema-usuario y los procedimientos que se aplicarán para verificar la integridad de la base de datos y la adecuación de la información contenida.

De acuerdo al esquema anterior, una forma de reducir el error durante la entrada de datos, y su posterior combinación y preparación para el análisis estadístico, consiste precisamente en intervenir sobre todos aquellos aspectos relacionados con la elaboración del diseño de la base que podrían incidir en un incremento del número de errores. En los posteriores apartados veremos cómo se puede afectar positivamente la calidad de los datos tomando esta perspectiva como punto de partida (Moritz et al., 1995).

Nuestra experiencia particular nos ha demostrado la necesidad de concretar y sistematizar aquellos aspectos de calidad relacionados con la gestión de bases de datos, aspectos poco considerados en los estudios de investigación que se llevan a cabo en nuestro país. Ello justifica sobradamente la inclusión de los próximos apartados en este proyecto. Únicamente destacaremos, como preámbulo, que un diseño adecuado de la base de datos supone que la información registrada en una investigación no contiene errores evitables automáticamente, que contiene un número de errores no detectables automáticamente inferior al que resultaría si la entrada de datos se efectuara sin un diseño previo o con un diseño no adecuado, y que contiene un menor número de valores desconocidos. Además, los datos que resultan de diseños correctos permiten obtener índices fiables sobre calidad de la gestión y sobre la eficacia/eficiencia de los operadores implicados en la grabación de la información.

Previamente al desarrollo del resto de los apartados que componen este capítulo, apuntamos unas breves notas históricas sobre la teoría relacional en el ámbito de bases de datos (Frost, 1989). En este punto, se considera que los trabajos de Edgar F. Codd en torno al álgebra relacional y de sus implicaciones en la configuración de bases de datos constituyen los

pilares fundamentales sobre los que se desarrolla la teoría relacional (Cood, 1972, 1979). El impacto de este autor y las aportaciones de sus coetáneos produjo, en la década de los 80, un notable desarrollo de los sistemas de bases de datos desde la óptica relacional (Date, 1989). Paralelamente, se producía una importante dispersión de las pautas que eran seguidas en los distintos trabajos. Por esta razón, en un intento de homogeneización de criterios, Cood propuso las reglas que debía seguir todo sistema de bases de datos para ser considerado como plenamente relacional (Cood, 1985a, 1985b). Aunque estos criterios fueron ampliamente aceptados desde su aparición, en la realidad se han tomado como puntos de referencia para que los diseñadores construyan sistemas relacionales, y todavía hoy no existe ningún sistema en el mercado que consiga cumplir todas las reglas propuestas por Cood.

### 3.3.3. Mejora de la calidad a través de la definición de la estructura de la base de datos

El diseño de la base de datos se inicia con la definición de la correspondiente estructura que subyace a la información. Si se adopta un modelo relacional, este proceso consiste en la aplicación iterativa de un conjunto de reglas, denominadas formas normales (Codd, 1985a, 1985b), cuyo objetivo consiste en detectar y eliminar progresivamente las redundancias, duplicidades<sup>5</sup> y valores nulos innecesarios. Evidentemente, este diseño contemplará qué aspectos de la realidad deben registrarse y la forma de hacerlo.

El proceso de normalización se inicia creando una tabla que contenga toda la información de interés para los objetivos de un estudio ("relación universal"). A continuación se definen nuevas tablas derivadas de esta tabla inicial que contienen una parte de los campos, a través de un proceso de descomposición sin pérdidas (esto es, la reunión de los archivos resultantes debe contener exactamente la misma información que el archivo original). Las diferentes tablas guardan relaciones a través de campos que posibilitan esta reunión (en este contexto se denominan claves de relación), y deben mantener una estricta dependencia funcional que garantice la integridad referencial.

La normalización afecta y mejora en diferentes modos la calidad de la gestión de datos de un estudio (Date, 1989). Por ejemplo, la primera forma normal garantiza que los registros aparezcan correctamente identificados (de forma unívoca), y también que la información que se plantea capturar para responder las hipótesis de un estudio (y sus relaciones) corresponda a la estructura de campos de la base. De este modo, la descomposición a través de la primera forma normal garantiza que no existan dos registros idénticos, que no existan grupos de repetición (campos de ocurrencias múltiples) y que los campos sean indivisibles. Asimismo, las formas cuarta y quinta garantizan la actualización de los datos (eliminando redundancias y duplicidades), que no se producirán errores durante el mantenimiento de la base de datos (inclusión, modificación y borrado de información), y que se podrán definir matrices rectangulares a partir del caso de estudio que posteriormente serán objeto de análisis estadísticos.

En palabras de Litton (1991), el diseño de la base de datos constituye un proceso iterativo cuyos objetivos fundamentales consisten en garantizar que: (1) cada objeto quede representado en un único registro, sin que pueda haber dos registros que contengan la misma información; (2) cada tabla se construya a partir de registros que representen objetos similares

---

<sup>5</sup> En el marco de trabajo de la teoría de bases de datos es habitual diferenciar entre información redundante y duplicada. La información redundante se podría extraer de otra información ya registrada en la base, por ejemplo a partir de una operación matemática. Este sería el caso de registrar el índice de masa corporal cuando se dispone del peso y la talla, o registrar la edad del sujeto cuando se dispone de la fecha de nacimiento y respuesta a un cuestionario. La información duplicada no se puede obtener directamente aplicando una operación a los datos. Por ejemplo, registrar el nombre de los trastornos CIE-10 cuando se dispone de un campo con los respectivos códigos numéricos.

del mundo real; y (3) al construir cada una de las tablas que forma parte del estudio se almacenará la información pertinente (sólo la información necesaria que está directamente relacionada con la temática de la tabla y con los objetivos del estudio), de la forma más adecuada, con el mínimo número de relaciones, asegurando la ausencia de duplicidades y redundancias, y garantizando la integridad de la base de datos respecto a las operaciones habituales de inserción, edición y borrado de registros.

Finalmente, destacar que durante la fase de diseño los SGBD relacionales suelen permitir implementar diferentes tipos de relaciones entre las tablas. Entre las dos más habituales destacan: (1) la relación "uno a varios", en la que un registro de la tabla padre puede tener más de un registro en la tabla hija, pero cada registro de la tabla hija sólo tiene un registro como máximo en la tabla padre; y (2) la relación "uno a uno", donde un registro de la tabla padre no puede tener más de un registro en la tabla hija, y un registro de la tabla hija no puede tener más de un registro coincidente en la tabla padre.

### **3.3.4. Mejora de la calidad durante la grabación de los datos**

Los elementos para garantizar la calidad de los datos tomando como referente su grabación consisten en garantizar que la información esté exenta de errores, que no contenga valores omitidos (desconocidos) y que se ajuste a los momentos temporales de registro establecidos en el diseño (especialmente en el caso de estudios longitudinales). Para ello es posible utilizar diferentes tipos de controles, los cuales pueden o no estar automatizados según el SGBD utilizado.

#### ***3.3.4.1. Controles automáticos de calidad: definición de restricciones***

Uno de los aspectos fundamentales que se aborda en la actualidad desde las teorías de bases de datos consiste en establecer y especificar las restricciones entre datos, esto es, en definir el conjunto de valores admisibles en cada campo y las correspondientes condiciones de edición. Estas restricciones se implementan a través de controles automáticos que expresan la consistencia de forma matemática (dando lugar a un error) o probabilística (dando lugar a un aviso). Se considera que estos sistemas mejoran la calidad de los datos en la medida en que constituyen protecciones que chequean, detectan e impiden los errores habituales de registro: datos incorrectos y datos ausentes que posteriormente imposibilitan el correcto diagnóstico de los síndromes.

Atendiendo al momento en que son evaluadas, es posible establecer tres tipos de restricciones (Losilla y Navarro, 1997): pre-validaciones, in-validaciones y post-validaciones. Las pre-validaciones constituyen condiciones lógicas asociadas a un campo que se evalúan antes de escribir ningún valor en dicho campo, y que determinan si se permite la edición habitual, o si se debe completar con un valor no aplicable, o si se le debe asignar un valor automáticamente (desconocido o distinto de desconocido). Las in-validaciones (en ocasiones denominadas máscaras o plantillas) garantizan que los valores grabados coincidan con el tipo de dato permitido en cada campo, ya que constituyen aquellas condiciones que se evalúan durante la entrada de cada valor en su correspondiente campo, y permiten establecer qué transformaciones globales se pueden realizar una vez escrito el valor (por ejemplo, cambio de unidad de medida). Finalmente, las post-validaciones determinan (de forma absoluta y/o relativa) qué rango de valores son admisibles en un campo y/o qué condiciones deben cumplir dichos datos, evaluando cada unidad de información después de que ésta haya sido introducida.

Asimismo, cuando se diseñan estudios y se utilizan sistemas de bases de carácter relacional, es importante tener presente que la calidad de la gestión sólo se garantiza si la secuencia de restricciones conlleva (Gassman et al., 1995): (1) chequear si cada variable

cumple las condiciones de consistencia que afectan a otras variables del mismo cuestionario; y (2) chequear si las variables cumplen condiciones de consistencia que afectan a varios cuestionarios (especialmente cuando se programan seguimientos).

Finalmente, también resulta imprescindible tener presente que en ocasiones se editan y se modifican las protecciones de las variables posteriormente al registro y grabación de los datos. En estos casos es fundamental documentar tanto las causas de la edición como el momento en el que se han realizado. Asimismo, es básico volver a chequear de forma retrospectiva los datos ya grabados (a ser posible de forma automática) para comprobar que son consistentes con las nuevas condiciones de edición.

Pese a las ventajas de prever chequeos informáticos a los datos del tipo antes expuesto, en ocasiones resultan insuficientes. Por ejemplo, algunos programas informáticos de gestión de datos no incorporan la posibilidad de definir este tipo de restricciones. En otras ocasiones, cuando la grabación de la información no se efectúa directamente (por ejemplo, la grabación magnética de la información se realiza en etapas posteriores al registro en lápiz y papel) los investigadores pueden considerar que los chequeos enunciados son insuficientes para garantizar realmente la calidad de grabación de datos. En estos casos es necesario recurrir a procedimientos alternativos para incrementar, en la medida de lo posible, la correcta grabación y registro de la información. Aunque en la actualidad se dispone de tecnología que facilita enormemente este tipo de tarea (Barton et al., 1991), por ejemplo los sistemas electrónicos de escaneo o las técnicas de reconocimiento de voz (aunque el desarrollo de estas últimas técnicas es muy prometedor, parece que todavía tardaremos un tiempo en disponer de productos fiables para utilizar en investigaciones), tradicionalmente los procedimientos para garantizar la calidad eran menos sofisticados y requerían, cuanto menos, un mayor esfuerzo por parte de los responsables de la gestión de datos. En el siguiente subapartado presenta los dos procedimientos básicos de control de calidad que todavía continúan siendo ampliamente utilizados: la comparación visual y la doble entrada. También se presenta una nueva técnica que está demostrando ser especialmente útil para mejorar la calidad de la captura de datos: la verificación aleatoria.

#### ***3.3.4.2. Controles no automáticos de calidad: comparación visual, doble entrada y verificación aleatoria***

Existen diferentes estrategias procedimentales no automáticas especialmente concebidas para mejorar la calidad del registro de los datos. Entre las más utilizadas destacan: la comparación visual y la doble entrada.

La comparación visual consiste en que uno o más sujetos cotejen los datos grabados con los registrados en las fuentes documentales originales. En este sentido, se considera que esta técnica supone una estrategia adecuada para verificar la calidad de dos fuentes diferentes: la de los datos originales y la de los datos grabados. A nivel procedimental, existen muchas variaciones de la comparación visual, dependiendo de quién realiza la comparación, del tipo de documento en el que están escritos los datos originales, del método utilizado para visualizar la información ya grabada, del lugar dónde se produce la comparación, del tiempo que transcurre entre la grabación y la comparación, y del número de comparaciones visuales que se efectúan.

Se considera que el mínimo grado de comparación visual consiste en el chequeo inmediato que realiza el operador que graba los datos en el momento en que registra cada valor; en este caso, el operador asume dos responsabilidades, la de grabación y la de verificación. De otro lado, la comparación visual diferida consiste en comparar datos grabados con las fuentes documentales originales una vez que la información ha sido entrada en la base de datos. Es habitual que cuando se utiliza la comparación visual diferida los datos se

impriman en formato similar al de las fuentes documentales originales para facilitar la contrastación.

Aunque hay factores importantes implicados en el éxito de la comparación visual (la forma en que se han transcrito los datos ya grabados o factores meramente ambientales), se ha postulado que la eficacia de la técnica depende, básicamente, de las características del sujeto que la lleva a término. En este sentido, los factores determinantes del éxito serían variables del tipo: grado de motivación, de fatiga, de estrés o de aburrimiento. Por esta razón, y aunque algunos consideran que la técnica de verificación visual es muy útil cuando no es posible adoptar otras estrategias más potentes como la doble entrada, se suele considerar que constituye una estrategia procedimental muy elemental que sólo es adecuada cuando se usa como complemento de otros métodos de control de calidad, cuando se dispone de pocos datos o cuando el sujeto que la lleva a término asume la responsabilidad de conducir otras fases de la investigación (como el análisis de datos) y las consecuencias de la presencia de errores (Blumenstein, 1993). En cualquier caso, se trata de una técnica costosa en cuanto al tiempo requerido para llevarla a término que no necesariamente produce mejoras considerables en la calidad de los datos. Por ello, los investigadores prefieren utilizar procedimientos tales como la doble entrada o la verificación aleatoria.

Como su nombre indica, la doble entrada de datos es una técnica de control que se aplica durante la fase de grabación de los datos, y que consiste en grabar dos veces la información. A nivel procedimental, la estrategia de la doble entrada se puede llevar a cabo por una única persona u operador de entrada de datos -encargado de entrar la misma serie en dos ocasiones diferentes- o por dos personas distintas -cada etapa de la grabación sería responsabilidad de un operador diferente-. Algunos autores sugieren que la doble entrada con dos operadores constituye el procedimiento de elección, especialmente cuando los datos que se deben transcribir y registrar están escritos a mano o en un protocolo que dificulta su comprensión, ya que en estos casos pueden aparecer diferentes problemas (por ejemplo, de legibilidad) y es importante disponer de dos interpretaciones independientes de los datos (Blumenstein, 1993; Cobos, 1995; Gassman et al., 1995). Tras el doble registro de los datos se procede a su comparación automática, que se puede realizar de forma inmediata o diferida. En el primer caso, el operador que registra la serie de datos en la segunda ocasión tiene la oportunidad de corregir las discrepancias originadas respecto a la entrada (aunque idealmente no se le debería permitir acceder a los datos que se grabaron la primera vez). En el caso de la comparación diferida, debe existir un tercer sujeto que revise la serie de datos en ambas ocasiones y resuelva las discrepancias, de forma que los operadores de entrada de datos están liberados de dicha responsabilidad. Además, la comparación diferida debería comportar la disponibilidad de un procedimiento de guía para los responsables de resolver las discrepancias. Asimismo, cuando se editan valores ya registrados para su corrección, es importante que los sistemas dispongan de la posibilidad de guardar los valores originales, los valores corregidos, la fecha de la modificación, las razones que han conducido a la corrección y las personas que las realizan y/o autorizan (Gassman et al., 1995).

Los estudios realizados en torno a la doble entrada constatan que esta técnica reduce sensiblemente el número de errores de registro y/o grabación respecto a la entrada simple (Blumenstein, 1993). En efecto, a pesar de las dificultades que supone la comparación de los resultados obtenidos en los diferentes estudios, fundamentalmente debido a que cada uno utiliza diseños y estrategias metodológicas diferentes, procedimientos de doble entrada distintos y operadores con diferentes niveles de motivación e implicación en el resto de la investigación y en el manejo de datos, los resultados obtenidos en diferentes estudios constatan que cuando se emplea esta estrategia de control la proporción de errores se reduce hasta en un 30% (Gibson, Harvey, Everett y Parmar, 1994; Wolf, 1993; Reynolds-Haertle y McBride, 1992), y también que la entrada simple comporta que el porcentaje de errores sea entre 4 y 5 veces superior (Neaton, Duchene, Svendsen y Wentworth, 1990). Estos resultados

han sido considerados como el principal argumento a favor del uso de dicha técnica en casi todos los ámbitos de investigación.

Desde un plano meramente funcional, algunos estudios también han evaluado el número y tipo de errores que se pueden detectar e impedir a través de la doble entrada. Así, por ejemplo, en un estudio llevado a cabo por Gibson et al. (1994) en el contexto de un ensayo clínico multicentro, se valoró la magnitud del decremento de los errores de grabación cuando se utilizaba la doble entrada de datos por dos operadores distintos. Uno de los atractivos de este estudio estriba en el hecho de que ambos operadores estaban entrenados en el uso de los formularios, asumían otras responsabilidades del estudio (además del registro de datos) y habían sido informados tanto de la importancia de un registro perfecto de la información como de las repercusiones que podía tener cada tipo de error dentro del contexto general de la investigación. Para ello se establecieron cuatro categorías de errores: muy importantes, importantes, poco importantes y triviales. Los autores de este estudio observaron que la disponibilidad de personal responsable encargado de la transcripción de los datos, entrenados en el uso de los formularios, y conocedores de los datos que deben ser grabados de forma perfecta, mejora notablemente la calidad de la información registrada y prácticamente hace innecesaria la doble entrada (aunque no se consiga reducir totalmente el error de registro en ninguna de las categorías descritas). Estas mismas hipótesis ya habían sido enunciadas en estudios anteriores (Stellman, 1989), y se han consolidado en investigaciones más recientes (Gassman et al., 1995), que señalan la conveniencia de que tanto el registro de los datos como el posterior manejo de la información sea efectuada en la "propia casa" (es decir, por los propios responsables del estudio), ya que la presencia de errores e inconsistencias se incrementa en relación directa con el uso de personal ajeno a los propósitos de la investigación (por ejemplo cuando se recurre a empresas comerciales o estudiantes).

Entre las mayores dificultades que se han relacionado con la doble entrada destaca el tiempo que requiere. Aunque pocos estudios han medido el coste real que supone reentrar la totalidad de los datos (Van Der Putten, Van Der Velden, Siers y Hamersma, 1987), algunos trabajos estiman que incrementa en más de un 37% el tiempo requerido para la grabación magnética de la información (Reynolds-Haertle y McBride, 1992). Esto implica que la doble entrada sólo puede ser empleada en escasos estudios que disponen de los suficientes recursos como para implementar mecanismos de control de datos tan costosos, por ejemplo en los ensayos clínicos. También se ha considerado que la doble entrada no garantiza totalmente la calidad del registro si los datos no son sometidos de forma simultánea a chequeos para detectar las inconsistencias que podrían contener las fuentes documentales originales (Cobos, 1995). Finalmente, la última dificultad relacionada con la doble entrada consiste en la imposibilidad de realizarla cuando los datos se graban directamente en la base de datos, por ejemplo cuando la entrevista se realiza de forma directa con el paciente y se utiliza el ordenador para introducir la información. En esta situación, la alternativa más adecuada a la doble entrada consistiría en el uso de chequeos inmediatos e intensivos del tipo antes expuesto (pre-validaciones, in-validaciones y post-validaciones), y la inclusión de preguntas redundantes que permitan confirmar la información más crucial (Blumenstein, 1993).

Debido al coste que supone la doble entrada de la información, algunos autores han propuesto sistemas en los cuales unas variables son verificadas de forma más intensa y exhaustiva que otras. Por ejemplo, Mullooly (1990) describe un algoritmo basado en un modelo de probabilidad que permite implementar procedimientos para el control de la calidad que obligan a grabar dos veces el 100% de los ítems considerados cruciales, y sólo una vez los ítems menos importantes. Neaton et al. (1990) también discuten la posibilidad de implementar una "verificación selectiva", especialmente en el caso de que los ítems importantes no puedan ser sometidos a chequeos de consistencia. Finalmente, otros sistemas incluyen procesos de verificación aleatoria (Doménech y Losilla, 1995), es decir técnicas de verificación de un porcentaje total de los datos escogidos al azar por el sistema informático y presentados con un

retardo suficiente para evitar efectos de recuerdo; esta técnica requiere necesariamente que sea el mismo operador quien reintroduzca los datos, aunque tiene la ventaja de proporcionar un "feed-back" inmediato de los errores que se van cometiendo.

Entre las alternativas propuestas en el párrafo anterior consideramos que la verificación aleatoria ofrece diversas ventajas. Así, por ejemplo, se ha demostrado que esta estrategia reduce considerablemente el coste de la entrada de datos manteniendo gran parte de las características positivas que se relacionan con la doble entrada. Además, este sistema hace posible la obtención de una estimación del porcentaje de errores y un índice de la aptitud y eficacia de los operadores. En un reciente estudio con 45 operadores de entrada de datos (Doménech, Losilla y Portell, en prensa), se demuestra que la verificación aleatoria aumenta la eficacia de la entrada de datos (sin afectar la eficiencia de la grabación) en comparación en situaciones de control que no incluyen procedimientos para obtener indicadores sobre la calidad de los datos que se registran. Asimismo, los resultados obtenidos sugieren que el efecto de "feed-back" de la verificación aleatoria actúa como un mecanismo explicativo de la reducción del error.

La verificación aleatoria se puede implementar de forma compatible con la doble entrada. En este sentido, se podría utilizar durante la entrada de datos y completar al final del estudio los registros no verificados de este modo con doble entrada, o bien en situaciones en las cuales los operadores requieran diferentes niveles de control. Este último aspecto nos parece, cuanto menos, especialmente relevante. Así, por ejemplo, en un estudio de Reynolds et al. (1992) los autores observaron una gran variabilidad entre los 46 operadores encargados de entrar los datos en el estudio. De otro lado, recordamos que Gibson et al. (1994), que obtuvieron un porcentaje de error mucho menor, atribuyeron gran parte de esta mejora en la grabación de los datos al uso de operadores (en este caso sólo 2) que habían sido entrenados en el uso de los formularios, que estaban responsabilizados de otros aspectos también relevantes de la investigación y que por ello estaban concienciados de la importancia de un registro correcto de los datos.

En cualquier caso, y a pesar de las ventajas que supone a nivel de reducción de costes, la verificación aleatoria no está exenta de algunas de las dificultades señaladas para la doble entrada. Así, por ejemplo, esta técnica tampoco garantiza la calidad del registro si los datos no son sometidos de forma simultánea a chequeos que detecten las inconsistencias de las fuentes documentales, ni tampoco suele ser posible aplicarla cuando la grabación de la información se realiza "on line".

Antes de finalizar este apartado nos gustaría destacar que las nuevas tecnologías disponibles simplifican el proceso de grabación de datos incidiendo en aspectos de control de calidad. Así, por ejemplo, una de las prestaciones más utilizadas es el escaneo de las fuentes documentales originales como medio para construir el archivo de datos. Por ejemplo, el programa Teleform permite generar formularios precodificados que una vez cumplimentados son leídos por un escáner e introducidos directamente a la base de datos con un alto grado de fiabilidad. A pesar de las enormes ventajas que conlleva este procedimiento, posee una importante dificultad antes ya indicada para la doble entrada y la verificación aleatoria: en el caso de que no sean chequeados los datos después de la grabación, pueden existir errores en las fuentes documentales originales que no sean detectados y que se traspasen a la matriz de datos final que será objeto de análisis estadístico.

#### ***3.3.4.3. Otros controles de calidad: previsión de valores desconocidos***

Finalmente, al hablar de la calidad del registro y grabación de la información, también queremos destacar una importante restricción que, cuando se implementa como parte del diseño de la base de datos, incrementa notablemente la calidad. Se trata de limitar la posibilidad de que los operadores de entrada de datos no registren ningún valor en campos que



cumplen las condiciones de edición, o lo que es igual, decrementar la posibilidad que dejen campos sin completar o con valor desconocido.

La restricción anterior constituye una mejora notable en la calidad de los datos al abordar directamente uno de los problemas más importantes con los que muchos investigadores se encuentran cuando trabajan en el ámbito de las ciencias humanas: la problemática de los valores desconocidos. Esta dificultad no es, en absoluto, reciente. Cuando se realiza una investigación en el ámbito clínico es habitual no disponer de todos los datos de los pacientes, ya que a veces es imposible recoger toda la información requerida para la correcta evaluación psicopatológica de los sujetos a pesar de los esfuerzos realizados por muchos clínicos-investigadores (múltiples observaciones de los sujetos, consulta con otros profesionales o con el resto del personal del centro, revisión de las historias clínicas, etc.). En los contextos habituales en que se encuentran numerosos clínicos e investigadores (entrevista directa con el paciente), esta limitación suele ser consecuencia de una de las siguientes situaciones (Fowler, 1993):

1. El sujeto se niega a dar información. En este sentido, podríamos hablar de sujetos en los cuales se evidencia una clara falta de cooperación, bien sea por un "rechazo activo" a ser entrevistados durante la sesión de evaluación, o bien por un "rechazo pasivo" manifestado a través de reiterados intentos por posponer la entrevista. En nuestra investigación hemos observado que el rechazo activo es característico en algunos niños pequeños que adoptan actitudes claramente negativistas y se niegan a colaborar. En cambio, los rechazos pasivos son más habituales en algunos padres que solicitan aplazamientos sistemáticos de las sesiones de entrevista, o bien en aquellos otros que nos dificultan ponernos en contacto con ellos para acordar el día y/o la hora de las sesiones.
2. El sujeto no conoce o no dispone de la información cuando se le solicita. En este sentido, podría ser que los problemas de memoria de los sujetos dificultaran el recuerdo de algunos datos requeridos en la evaluación.
3. Se produce una interferencia entre diferentes síntomas o grupos de síntomas. Imaginemos un caso típico: un sujeto que presenta una determinada sintomatología depresiva pero es incapaz de dar respuestas inteligibles debido a que también presenta problemas severos de comunicación.
4. Finalmente, también es posible que el entrevistador posea dificultades para conducir adecuadamente una entrevista (por ejemplo, que no sea capaz de seguir las rutinas adecuadas debido a la falta de entrenamiento o a la complejidad del protocolo de evaluación, o que no alcance a registrar de forma adecuada la información que le facilita el entrevistado). Referir esta fuente de valores desconocidos en último lugar no debe, en absoluto, restarle importancia. De algún modo, permite que el investigador tome conciencia de que la responsabilidad de que aparezcan valores desconocidos no recae totalmente en el entrevistado, sino que también es importante que el entrevistador posea las habilidades necesarias y que el estudio esté correctamente diseñado.

La solución al problema de los datos incompletos ha sido, y sigue siendo, difícil. Tras revisar los trabajos publicados durante los últimos años, no hemos encontrado ninguna estrategia para la previsión y el tratamiento de los valores desconocidos que goce de consenso y aceptación general. Esta falta de acuerdo podría ser debida al hecho de que en muchos estudios con pacientes psiquiátricos los ítems con valores desconocidos no se distribuyen aleatoriamente, y la relación entre el número de campos no contestados y la severidad de diferentes aspectos de la psicopatología suele presentar relaciones lineales (Taylor y Amir, 1994). En relación a este punto, algunos investigadores consideran que no se debe realizar ninguna previsión previa de los valores desconocidos, ni tampoco intentar corregirlos, ya que

el patrón de distribución de los campos vacíos constituye un objeto de estudio más dentro de la investigación. En estos trabajos se argumenta que la configuración que presenta el patrón de valores desconocidos y sus correlatos con otros factores psicopatológicos ofrece una importante información práctica y clínica; por ejemplo, este patrón podría evidenciar que diferentes subgrupos de sujetos con una misma etiqueta diagnóstica en realidad corresponden a diferentes formas patofisiológicas. Otro grupo de trabajos matiza que la distribución de los valores desconocidos a lo largo de cada ítem, o de cada agrupación de ítems, provee a los investigadores de información muy relevante sobre la calidad del proceso y/o manejo de los datos (Stellman, 1989), especialmente de todos aquellos aspectos de registro y edición de la información.

Sea cual sea la postura que se adopte, parece existir una impresión bastante general entre clínicos e investigadores respecto a la indeseabilidad de que los datos contengan campos con valor desconocido. Sin embargo, tal como se confirma en una revisión realizada por Vach y Blettner (1991), la mayoría de los profesionales raramente explican cómo han abordado esta cuestión (si es que lo han hecho de alguna forma). Las pocas propuestas prácticas halladas en la literatura no suelen efectuar previsiones desde el punto de vista del diseño de la base de datos y la grabación de la información, sino que se limitan a ofrecer propuestas de estrategias para ser adoptadas en una fase posterior a la recogida de datos (habitualmente durante la preparación de la matriz para el análisis estadístico).

En contadas ocasiones, se presentan estrategias de prevención de valores desconocidos que se aplican de forma anterior a la creación de la matriz rectangular para el análisis estadístico, y que constituyen alternativas aplicables durante la recogida y grabación de los datos. En este sentido, los profesionales las consideran "soluciones" menos radicales. La base del razonamiento radica en el presupuesto de que la ausencia de información puede el diagnóstico, pero no por ello resulta un impedimento mayor que la imposibilita siempre. En estos casos, el clínico puede considerar válido su propio juicio y registrar sus impresiones, siguiendo diferentes criterios: (1) cuando dispone de la suficiente información, el clínico-investigador puede considerar que la falta de unos pocos datos no es justificación suficiente para impedir el diagnóstico, y por lo tanto puede verse tentado a registrar su propio juicio sobre una determinada respuesta en base al resto de la información disponible o en base a la impresión general que se ha forjado sobre el caso; y (2) cuando a través de otros medios diferentes al propio instrumento de evaluación se pueden recabar datos, el profesional puede optar por realizar inferencias sobre los valores que podrían ser asignados (por ejemplo, si un clínico o investigador observa que un paciente permanece mirando fijamente un punto de su entorno o que está hablando solo, puede llegar a suponer que tiene alucinaciones). Sin embargo, las dos estrategias anteriores poseen una importante limitación: se trata de diagnósticos clínicos que se realizan a partir de información que no es del todo fiable y/o válida, lo cual conlleva necesariamente cuestionar la confianza que merecen los resultados.

En la actualidad, al menos a un nivel teórico, comenzamos a encontrar en la literatura especializada autores como Taylor y Amir (1994) que sugieren el uso de estrategias secuenciales más formales para minimizar los problemas debidos a la falta de información. Estas estrategias comenzarían a aplicarse durante la fase de recogida de datos seleccionando instrumentos de evaluación que contemplen la posibilidad de que algunos valores puedan ser desconocidos, por ejemplo protocolos que dispongan de una opción de registro y tratamiento para campos vacíos. Estas estrategias secuenciales conllevan una notable mejora respecto a los procedimientos descritos anteriormente. Cuanto menos, denotan una mayor sensibilización por parte de los investigadores, que ya planifican "a priori" la existencia de dichos valores (durante la fase de diseño del estudio) y les asignan un código de registro. Se debe tener presente que cuando no se realiza esta previsión y se usan protocolos que no disponen de este código especial, se obliga al entrevistador para que tome una decisión en ausencia de respuesta: el ítem está presente (en un determinado grado) o ausente. Como es de suponer, a

pesar de que el formato de estos protocolos obligue a una "elección forzada", muchos investigadores optan por dejar en blanco estos campos. Pero cualquiera de las soluciones anteriores, dejar en blanco el campo cuando no se dispone de un código de valor desconocido o registrar un valor sin la suficiente información, es problemática. Por ejemplo, si el investigador deja en blanco los campos está transformando la escala e incidiendo sobre las propiedades psicométricas, sobre todo cuando se obtienen puntuaciones totales a partir de sumas de ítems (en este caso no suele existir ninguna indicación sobre cómo agregar los campos que no disponen de valor). De otro lado, cuando el entrevistador tiene que tomar la decisión sobre la presencia o la severidad de la psicopatología, los elementos que tendrá en cuenta en el momento de realizar esta elección difieren entre los diferentes centros, e incluso entre los diferentes profesionales. Por ejemplo, algunos clínicos o investigadores pueden considerar que cuando el resto de signos y síntomas están presentes o son severos, los ítems vacíos se deben registrar en la misma línea. Pero otros profesionales pueden interpretar que cuando el sujeto no informa de un contenido es porque está ausente. Y aún más, algunos clínicos interpretan cualquier contenido no verbal a la hora de registrar una conducta o síntoma como presente (si un paciente habla solo indica que está alucinando), mientras que otros requieren mucha más información (disponer de documentación clínica sobre la presencia de alucinaciones anteriormente). Estas diferencias a la hora de tratar la información poseen consecuencias negativas sobre la fiabilidad de los instrumentos.

En síntesis, por lo expuesto hasta el momento parece sumamente relevante que todo clínico e investigador reconozca, prevea y asuma la posibilidad de que se produzcan ítems con valor desconocido, de forma que se puedan adoptar las mejores soluciones desde las primeras etapas del diseño de la investigación: Si la calidad de datos se enfoca formalmente desde la teoría del diseño de bases de datos, y se conocen las dificultades de tratamiento y manejo de información desconocida, es sumamente interesante adoptar restricciones que limiten la posibilidad de que cualquier operador de entrada de datos pueda dejar campos vacíos cuando éstos cumplen las condiciones de edición. En el capítulo en el que se presenta el proceso de informatización de la entrevista DICA-IV mostraremos qué restricción particular hemos adoptado en relación a esta problemática: limitar la posibilidad de que los sujetos dejen campos que cumplen las condiciones de edición sin grabar (esto es, vacíos).

### **3.3.5. Mejora de la calidad a través del diseño de la interfaz**

Diversos autores plantean una clasificación general de los tipos de interfaz de entrada de datos de acuerdo a tres criterios diferentes (Fowler, 1993; Kinnear y Taylor, 1991; Saris, 1991): tipo de administración (autoadministrada "versus" heteroadministrada), tipo de periférico (manual "versus" automático) y fuente del dato o registro (no informática "versus" propio sujeto). Atendiendo a esta clasificación, es posible determinar el espacio potencial de fuentes de error que no puede ser controlado de forma directa a través del diseño de la base de datos propuesto anteriormente. En este conjunto de errores potenciales destaca toda aquella información que cumple con las restricciones especificadas con anterioridad, pero que no coincide con el valor original informado por los sujetos evaluados o representativo de la realidad que se pretende medir/registrar (Karwowski, Eberts, Salvendy y Noland, 1994; C.G. Wolf, 1992). Los trabajos actuales demuestran que se pueden establecer ayudas a través del interfaz del registro informatizado con objeto de reducir los errores que no pueden ser detectados de forma automática, recurriendo a los resultados que se desprenden de los trabajos de la HCI o de la construcción de reactivos para entrevistas y cuestionarios. En apartados anteriores ya hemos sugerido que el diseño adecuado de los instrumentos de evaluación es fundamental de cara a mejorar la calidad de la gestión de datos, y que los déficit del diseño contribuyen a incrementar, entre otros, los errores de medida (Sánchez, 1992). Así, por ejemplo, el uso de indicaciones no verbales (por ejemplo gráficas) y la formulación correcta

de las preguntas pueden mejorar o reducir notablemente la calidad final de los datos, ya que ayudan a transmitir los objetivos y la relevancia de cada ítem. Curiosamente, los aspectos de diseño sólo se suelen tener en cuenta cuando se construyen cuestionarios para ser autoadministrados. Sin embargo, contrariamente a lo que algunos todavía puedan suponer, este aspecto siempre debería tomarse en consideración, ya que según muestran recientes estudios ni tan sólo un alto grado de experiencia de los entrevistadores en un determinado formulario puede suplir los déficit de diseño. En el siguiente apartado aportamos una breve introducción a los trabajos aparecidos dentro del ámbito de la HCI en relación a la calidad de datos científicos.

### ***3.3.5.1. Los trabajos aportados desde el ámbito de la HCI***

Como su nombre indica, la "Human Computer Interaction" trata del diseño, análisis y valoración de sistemas informáticos interactivos. Su principal objetivo es identificar factores o variables que influyan significativamente en la interacción entre hombre-ordenador, y en derivar empíricamente métodos y guías que ayuden a construir sistemas informáticos (Lowgren y Lauren, 1993; Hartson y Hix, 1989). Para ello ha sido habitual la confrontación de dos tipos fundamentales de sistemas interactivos: los centrados en la tarea (cuyo objetivo es ofrecer al usuario todas las funciones necesarias para realizar las tareas de forma adecuada y flexible) y los orientados al usuario (que intentan minimizar el problema de la interacción).

Durante mucho tiempo, el diseño de interfaces ha sido el principal objeto de estudio de la HCI. Lo importante en este sentido no ha sido la definición de interfaces en términos descriptivos, sino en términos normativos: la HCI tiene como objeto último proponer cómo deberían ser los modos de interacción para asegurar que los sistemas maximicen su "usabilidad". Para alcanzar este propósito, un volumen considerable de trabajos han evaluado las prestaciones que ofrecen diferentes factores tales como las formas de interacción (menús, comandos, etc.), los mecanismos de "input" (teclado, ratón, pantallas táctiles, uso de sonidos, etc.), los sistemas de ayuda disponible ("on-line", tutoriales, escritos, etc.), los diferentes modos de interacción de los usuarios finales (por ejemplo en relación a la experiencia previa con los sistemas), el diseño de las pantallas (forma de presentar la información, nivel de iluminación, etc.), el grado de funcionalidad de los sistemas (qué tareas recaen en el usuario y cuáles en el programa), etc.

El interés por el estudio de la interacción hombre-ordenador aparece en el momento en que se dispone de sistemas informáticos cuyo manejo requiere un considerable esfuerzo por parte de los seres humanos, sobre todo de tipo cognitivo. Pensemos que hasta la década de los 70, muchos diseñadores de software solían ser a su vez los usuarios, de manera que no existía una preocupación importante por facilitar la tarea a otros usuarios potenciales (simplemente, éstos se adaptaban a la máquina). Sin embargo, a partir de la década de los 70, la rápida renovación de las tecnologías informáticas y su difusión a poblaciones con diferentes niveles de experiencia, obliga a los diseñadores a tener presente que los usuarios finales deben "re-aprender" continuamente cómo funcionan los nuevos programas, y también que es preciso evaluar las capacidades de las diferentes poblaciones de potenciales usuarios (Gaines y Shaw, 1986a, 1986b; Van der Veer, 1989). En particular, durante estos años se evidencia la necesidad de evaluar las formas en que los usuarios perciben la máquina, diferenciando claramente entre programadores de software y sujetos con poca experiencia en informática. También se hace evidente la necesidad de evaluar las dificultades en el manejo de los sistemas, establecer en qué grado los usuarios cometen errores predecibles, y las formas de evitarlos (Booth, 1991; Rath y Brown, 1995). En síntesis, entre los objetivos fundamentales de los diseñadores del nuevo software destaca la posibilidad de conseguir que los usuarios realicen mejor y más rápidamente sus tareas. Para llevar a cabo este objetivo, en un principio primaron las consideraciones ergonómicas sobre las cognitivas o de diálogo. En la década de

los 80, sin embargo, se incrementa el número de investigaciones que evalúan las posibilidades que ofrecen los ordenadores personales interactivos con pantallas gráficas de alta calidad, el uso de modelos conceptuales familiares para el usuario, la adopción de comandos universales simples (capaces de realizar las mismas funciones en diferentes programas), la personalización por parte del usuario (esto es, la posibilidad de que cada usuario elija sus preferencias), etc. En los últimos años, la HCI ha pasado a ser considerada como una disciplina relativamente estable, y un gran número de compañías mantienen laboratorios para ayudar a desarrollar y evaluar software y hardware que facilite las tareas de los usuarios. La lectura de muchos de los trabajos recientes desarrollados en este campo permiten concluir que gran parte de los investigadores han focalizado su atención en el estudio de la acción humana (González, 1993). En este sentido, la acción se ha considerado como un elemento integral dentro de un sistema jerárquico de objetivos y subobjetivos (sistema manejado en términos de modelos de interpretación o representación mental de las acciones y de las organizaciones estructurales). Esta aproximación implica un cierto descuido del objetivo general de estudio de la relación entre el objeto tal como el usuario se lo representa mentalmente y el objeto tal y como el ordenador lo presenta. Los diseñadores de software, que cada vez están más concienciados de la distancia entre el objeto ya elaborado y la representación mental que el usuario se hace de tal objeto, han trabajado en la creación de productos que satisfagan y garanticen el concepto de "lo que los usuarios ven sea lo que verdaderamente consigan" ("what you see is what you get"). Este objetivo ha producido resultados muy notables (Eason, 1991), como por ejemplo la transformación del estilo de interfaz basado en comandos al estilo de interfaz basado en gráficos (y probablemente, en un futuro que no se augura demasiado lejano, un estilo basado en realidad virtual). Actualmente, los investigadores interesados en esta disciplina sostienen que los trabajos aparecidos en el seno de la HCI constituyen medios para obtener conocimientos y metodologías que ayuden a la creación de objetos en distintas disciplinas, entre ellas la psicología.

A pesar de las enormes facilidades conseguidas durante los últimos años en el diseño de interfaces, los problemas de interacción hombre-ordenador no están, en absoluto, resueltos. A la hora de abordar las dificultades que todavía quedan por resolver se pueden adoptar y combinar diferentes puntos de mira: (1) la de los tecnólogos y gran parte de los diseñadores de "software", que abogan por hallar soluciones en los avances que nacen de la tecnología informática; y la (2) la de clínicos e investigadores, que abogan por atender y examinar rigurosamente las necesidades y las expectativas de los usuarios finales con objeto de elaborar sistemas más prácticos y fáciles de manejar (esta última postura es propia de la psicología). La resolución de las dificultades planteadas bajo cada una de estas perspectivas obliga a concebir la HCI como una disciplina eminentemente interdisciplinar (Dix, Finlay, Abowd y Beale, 1993; Norman y Thomas, 1991). En efecto, los trabajos que aparecen en este ámbito recogen propuestas surgidas, por ejemplo, del seno de la psicología cognitiva (que proporciona conocimiento de las capacidades perceptivas, cognitivas y de solución de problemas), de la psicología experimental (que aporta procedimientos para evaluar y mejorar la "usabilidad" de los sistemas), de la ergonomía (para abordar cuestiones físicas), de la sociología (que ayuda a delimitar el contexto más amplio de la interacción), o de la propia ingeniería y/o tecnología informática (que posibilitan el desarrollo tecnológico necesario). En particular, la investigación psicológica desarrollada dentro del ámbito de la HCI se ha centrado en diferentes aspectos: en el estudio de tareas "globales" (por ejemplo, el análisis del comportamiento y razonamiento organizacional) y "locales" (el manejo y operación de un programa), en el estudio de los tiempos de ejecución de operaciones que implican el uso de ordenadores y en el estudio de la organización estructural de la tarea y su relación con los procesos de representación mental.

Además del carácter multidisciplinar, los problemas abordados desde la HCI se han contemplado desde diferentes posturas teóricas. Entre los marcos de trabajo más reconocidos

destacan (González, 1993): (1) los modelos de representación mental o conceptual (como el de GOMS), que constituyen una aplicación de los trabajos de Newell y Simon (1972) y proponen la necesidad de generar conocimientos generalizables que los diseñadores puedan utilizar, a partir de la definición y evaluación de los objetivos del usuario y la estructura que impone la tarea para su descomposición y resolución; (2) las teorías que valoran el contexto como objeto, lo cual permite valorar de manera realista la usabilidad de los sistemas y las interfaces; (3) las teorías que valoran el proceso como objeto, considerando especialmente las necesidades de los usuarios, sus formas de trabajo, la aceptación que realizan de los dispositivos de los sistemas, etc.; y (4) las teorías que valoran el sistema como objeto, observando los programas y prestaciones existentes con objeto de mejorarlos, por ejemplo construyendo bases de datos acerca de los éxitos y/o fracasos de los productos en fase de investigación y de los ya desarrollados.

Sea cual sea la postura teórica que se adopte, parece existir una cierta unanimidad respecto a la filosofía que subyace en el diseño de las interfaces. En este sentido, el objetivo fundamental es permitir al usuario realizar una tarea y estructurar su realización (en cuanto al número y tipo de operaciones que resulten las más adecuadas a las capacidades cognoscitivas y ergonómicas de los potenciales usuarios no expertos). La estructuración se puede realizar, por ejemplo, en términos de lenguajes y elementos simbólicos-conceptuales que posteriormente servirán para formular las instrucciones que permitirán traducir los objetivos-intenciones del usuario al sistema informático (Van der Veer, 1989).

Asimismo, también parece ampliamente reconocida la participación de factores semánticos, pragmáticos y significacionales en la interacción hombre-ordenador. Estos factores ayudarían a determinar el tipo de errores cometidos durante la ejecución de una tarea, cómo se organiza estructuralmente la tarea, y cuál es su representación mental. En particular, los factores semánticos estarían estrechamente relacionados con la elaboración de los modelos mentales que el usuario utiliza para aprender y comprender el funcionamiento del sistema, y para desarrollar las correspondientes tareas. Y los factores significacionales aparecerían relacionados con las intenciones que el usuario sostiene acerca de la tarea y del resultado que espera del ordenador. Precisamente, se ha hipotetizado que una de las mayores dificultades inherentes a la interacción hombre-ordenador radica en la falta de correspondencia semántica entre cómo el usuario piensa llevar a cabo la tarea y cómo impone el ordenador que la tarea sea realmente efectuada. Esta distancia sería consecuencia del grado acuerdo/desacuerdo entre el lenguaje en el que los usuarios quieren formular las demandas (de acuerdo al correspondiente modelo mental del que disponen), de las operaciones implicadas en su realización y del lenguaje que el ordenador pone a su disposición o les impone.

En la actualidad, también existe un amplio consenso en relación a las fases que conlleva el diseño de una interfaz: (1) el análisis de requisitos y tareas (lo cual implica la evaluación del contexto de trabajo y las necesidades de los usuarios finales); (2) la especificación y valoración de los diseños previos existentes que contemplen las necesidades del contexto y del usuario; (3) el desarrollo y la evaluación de prototipos; (4) el desarrollo del interfaz de usuario; y (5): la evaluación empírica a través de pruebas de usabilidad.

Una característica de los estudios realizados en el ámbito de la HCI radica en su inespecificidad respecto al tipo de aplicación informática, y también en su generabilidad respecto a los objetivos que se persiguen (por ejemplo, eficacia, eficiencia, satisfacción, motivación, etc.). Dentro de esta disciplina, gran parte de los trabajos sobre diseño del interfaz que contemplan la calidad de datos como objeto principal de estudio, contemplan el hecho de que los diferentes usuarios potenciales difieren en múltiples aspectos, tales como las habilidades que poseen en informática, el grado de entrenamiento, sus estilos cognitivos, sus habilidades visuales, su nivel de motivación, etc. Por esta razón, muchos trabajos han tratado aspectos tales como los sistemas gráficos orientados a la tarea, la disposición y el contenido de los menús, los tipos de controles, los mecanismos que el usuario puede utilizar para indicar las

acciones que a continuación realizará, y los efectos que los estados fisiológicos y contextuales ejercen sobre el grado de eficacia y satisfacción de los operadores (Doménech et al., 1997; Gao et al., 1990; Henning, Sauter, Salvendy y Krieg, 1989; Karwowski et al., 1994; Waersted, Bjorklund y Westgaard, 1991; C. G. Wolf, 1992). Todos estos componentes se relacionan con uno de los conceptos básicos estudiados bajo el seno de la HCI, que más estrechamente se ha relacionado con la calidad de datos: el de "usabilidad", esto es, la habilidad del usuario para obtener una representación del problema que comporta el uso de un programa de ordenador y el grado de eficacia que aporta dicha representación. Este concepto ha resultado de notable interés, debido fundamentalmente a la necesidad de que múltiples usuarios con distintas habilidades en informática utilicen los mismos programas con los mismos interfaces. Entre los componentes que incluye dicho concepto, la literatura matiza los más significativos (Dillon, 1983; Werner, 1996): el correcto diseño del entorno de trabajo y de los aspectos físicos de la interfaz, la facilidad de uso del sistema (idealmente, los sistemas deberían actuar de forma "controlable" y "transparente" para el usuario), la facilidad de aprendizaje (facilitada, por ejemplo, por una buena documentación técnica), la fiabilidad del sistema (los usuarios no deberían preocuparse de que un error atentara de forma catastrófica contra la integridad de los programas), la reducción en la frecuencia de errores, la facilidad de resolución y manejo de los errores cuando éstos se presenten, la adecuación de los sistemas de comunicación entre el usuario y la máquina, y la rapidez del sistema para facilitar resultados. Algunos autores establecen que un buen diseño orientado a la "usabilidad" debería conseguir una adecuada armonía e interacción entre 4 componentes básicos: el usuario, la tarea, el sistema y el entorno (Bannert y Kunkel, 1991).

La filosofía que subyace al concepto de "usabilidad" consiste en un intento de realizar tareas a través de la máquina de forma similar a los modos naturales de comprensión y/o de razonamiento humano (Batra y Srinivasan, 1992). En definitiva, se trata de conseguir una "máquina amigable" ("the friendly machine") con interfaces que incorporen estilos de diálogo en lenguaje natural, cuya actuación sea contexto-dependiente, consistente y adaptativa. Al referirse a estos componentes, autores como Senay (1992) hablan de interfaces "inteligentes" capaces de interpretar de forma flexible los lenguajes de comunicación hombre-ordenador, y subrayan el hecho de que la disponibilidad de tales sistemas disminuye los tiempos de ejecución, incrementa la satisfacción de los usuarios y reduce la probabilidad de error en las interpretaciones y acciones que emprenden (Booth, 1991). En la medida en que se consigan estos objetivos, algunos autores proponen que el diagnóstico psicológico constituya un producto más cuyo desarrollo se adapte a los métodos y principios formulados en el seno de la HCI (Bannert y Kunkel, 1991).

Finalmente, destacar algunas de las consideraciones más relevantes que se han realizado desde el seno de la HCI en relación al diagnóstico psicológico informatizado y los elementos de mejora de la calidad de los datos. En primer lugar, parece evidente las ventajas de llegar a elaborar y disponer de una interfaz uniforme que integrara diferentes aplicaciones (fundamentalmente, el sistema de gestión de bases de datos, el programa de estadística y el procesador de textos), que utilizara funciones u operaciones generales entre diferentes aplicaciones, y que contemplara las necesidades específicas de cada elemento del proceso de diagnóstico. Estos elementos ayudarían a reducir el esfuerzo requerido de aprendizaje por parte de los usuarios. En segundo lugar, cuando se trabaja con sistemas de gestión de bases de datos resultaría especialmente adecuado que las interfaces contemplaran la posibilidad de definir de forma simple las "vistas" de las estructuras entre los datos, y también las consultas que definen reuniones de datos o transformaciones complejas que requieran recorrer la estructura relacional. En tercer lugar, el desarrollo de sistemas informatizados debería valorar concienzudamente todos aquellos aspectos del diseño de la interfaz que se relacionan con el concepto de "usabilidad". Este último punto implica que el diseñador del "software" (y, en el caso de que fuera necesario, del "hardware") posea una visión general del proceso de

diagnóstico completo, que comprenda los requisitos específicos de cada elemento del proceso, que realice los diseños en estrecha colaboración con clínicos e investigadores, que elabore "proyectos piloto" sobre los que realizar pruebas de laboratorio y estudios de campo, y que ofrezca todos los medios de ayuda posibles a los futuros usuarios (entrenamientos intensivos, manuales de referencia, tutoriales, etc.). Los estudios empíricos demuestran que cuando se contemplan estos elementos y se desarrollan diseños centrados tanto en el usuario como en las tareas que deben llevar a cabo, la actitud de los usuarios finales es más favorable, se reduce el tiempo requerido de entrenamiento y se ofrecen menos oportunidades para entrar errores en los datos (Bannert y Kunkel, 1991).

### **3.3.6. Mejora de la calidad durante la fase de preparación de la matriz de datos para ser analizada: creación y modificación de variables**

Otro de los aspectos que no ha recabado la atención que merece en el entorno de la gestión de datos clínicos es la fase de preparación de la matriz de datos para ser sometida a análisis estadísticos, especialmente en lo que se refiere a las transformaciones de la información (creación y modificación de variables).

Todo profesional debe tener presente que los datos de un estudio deben estar libres de errores antes de proceder a generar las variables necesarias para contrastar la/s hipótesis planteada/s. Por lo expuesto hasta el momento, los datos "originales" estarán libres de errores cuando se capturan de forma correcta, es decir, cuando se introducen en una base de datos con campos protegidos con todas las validaciones posibles. Sin embargo, disponer de valores originales correctos no garantiza que los resultados obtenidos en los análisis estadísticos sean también válidos, ya que habitualmente las variables que se utilizan en dichos análisis suelen ser variables "generadas" o "calculadas". Esto supone que la fase de preparación de la matriz de datos para ser analizada debe recabar tanta atención como cualquier otra fase relacionada con la gestión de datos, especialmente en los aspectos referidos a la modificación y creación de variables. Las estrategias que se pueden adoptar para mejorar la calidad de la gestión en esta fase del estudio son, fundamentalmente, dos: (1) conocer profundamente el funcionamiento de los comandos de creación y transformación de variables del software utilizado; y (2) prestar una especial atención al manejo de campos vacíos en las operaciones de transformación.

En este trabajo no abordaremos el primero de los dos puntos anteriores debido a que los comandos de creación y modificación de variables tienen un funcionamiento particular según el programa que se utilice para gestionar y analizar los datos. Sin embargo, sí que retomaremos la problemática que subyace al manejo de variables cuyo campo de registro aparece vacío, y en particular la dificultad que plantea el manejo de valores desconocidos.

#### **Manejo de valores desconocidos**

La previsión de valores desconocidos (por ejemplo asignando un código de registro especial o limitando la posibilidad de que se olviden campos sin grabar) no es la solución definitiva al problema. Esta actitud no mejora demasiado la calidad de la gestión de datos si no se acompaña de un manejo adecuado de los campos vacíos durante la preparación de la matriz de datos para el análisis estadístico. En este caso, las mayores dificultades se plantean en el momento de decidir si es oportuno estimar los valores desconocidos y qué procedimientos de estimación deben ser adoptados, y también cuando se definen los algoritmos de corrección de las pruebas para definir síntomas y diagnósticos.

Lamentablemente, pocas veces se tiene en cuenta el problema de los valores desconocidos cuando manejan y analizan los datos. Muchos trabajos se limitan a informar que han eliminado del análisis estadísticos todos los sujetos en los cuales algún predictor o



respuesta aparece vacío, aquellos casos que no alcanzan un criterio-porcentaje de variables completas, o aquellas variables que acumulan un gran número de valores desconocidos. Uno de los argumentos que utilizan los usuarios de esta estrategia consiste en suponer que tras la eliminación de los valores desconocidos el análisis estadístico se efectuará sólo con "datos correctos". Sin embargo, la realidad es que estos investigadores desconocen las implicaciones teóricas y prácticas que conlleva la exclusión directa de sujetos con valores desconocidos. De entrada, esta solución afecta la calidad de la investigación a nivel de validez interna y externa, ya que eliminar registros para los que no disponemos de información implica asumir (entre otros aspectos) un supuesto que raramente se cumple en investigación clínica: que los valores desconocidos se distribuyen de forma aleatoria. En este sentido, se debe tener presente que en gran parte de estudios que se realizan en Ciencias de la Salud, eliminar directamente los sujetos con valores desconocidos conlleva una pérdida importante del tamaño de la muestra estudiada y de su representatividad -especialmente cuando los sujetos eliminados poseen características que se relacionan tanto con los factores de riesgo como con las respuestas- (Vach y Blettner, 1991), sesgo que afecta notablemente tanto la potencia de los análisis estadísticos como la posibilidad de generalizar los resultados.

Otra estrategia bastante utilizada consiste en no eliminar al inicio todos los sujetos con algún valor desconocido, sino en ir incluyendo en cada análisis particular "todos los casos posibles". Esta solución no se puede considerar mucho más brillante que la anterior, ya que en realidad conduce a estudiar múltiples muestras solapadas en las que no siempre convergen los mismos individuos. Esto dificulta aún más la posterior interpretación de los resultados.

La consideración de las limitaciones planteadas anteriormente sugiere la necesidad de adoptar mejores estrategias para la correcta gestión de los valores desconocidos en la fase de preparación de la matriz de datos. En este sentido, por ejemplo, se ha sugerido la posibilidad de utilizar estrategias de estimación (tanto simples como complejas) para corregir la presencia de dichos valores. Como ejemplo de estimación simple tenemos el prorrateo; como ejemplo de estimación compleja podemos hablar de procesos de captura-recaptura similares al muestreo en dos etapas (Hook y Regal, 1992), del modelado estadístico (Espeland et al., 1992; Johnson, George, Shahane y Fuchs, 1992; Wei y Tanner, 1991) -en particular, las técnicas de imputación múltiples (Efron, 1994; Rubin, 1987, 1996; Rubin y Schenker, 1991)- o del uso de sistemas más sofisticados tipo red neuronal (Navarro, 1998). Lamentablemente, cada una de las estrategias anteriores tiene serias limitaciones (Crawford, Tennstedt y McKinlay, 1995; Greenland y Finkle, 1995), ya que cualquier procedimiento adoptado para limitar la presencia de los valores desconocidos conlleva la evaluación de la tipología de distribuciones que subyace. Por ejemplo, entre las dificultades del prorrateo destacan: (1) en primer lugar requiere que la distribución de los campos vacíos sea aleatoria y que las relaciones entre variables sean lineales; (2) en segundo lugar, cuando todos los ítems de una escala tienen valores desconocidos dicha asignación crea una escala "artificial", lo cual sugiere que este procedimiento sólo se puede utilizar cuando la mayoría de los ítems de una escala tienen valores válidos (Kochhar, 1991). Y entre las dificultades de los procedimientos más complejos basta señalar que muchas de estas estrategias no están directamente disponibles en los sistemas de análisis de datos habituales, y que no siempre se dispone de la suficiente evidencia empírica sobre su fiabilidad y/o validez.

Pasemos ahora a la etapa de creación y modificación de variables. En psicopatología los investigadores no suelen interpretar directamente los ítems de los protocolos que utilizan para la recogida de datos, sino que construyen indicadores posteriores de salud y comportamiento combinando las puntuaciones obtenidas por los sujetos en agrupaciones de variables. Lamentablemente, en la práctica, no todos los protocolos que utilizan los clínicos y los investigadores incorporan algoritmos de corrección para que puedan ser adoptados directamente por los usuarios, y también es habitual que los programas de análisis de datos sean rígidos en la forma en que manejan la información (algunos pueden ser extremadamente

sensibles a la falta de algunos datos, y otros no consideran en absoluto esta posibilidad). Esto implica que cada investigador (independientemente de sus conocimientos en diseño y/o metodología) asume la responsabilidad de la manipulación de los datos y la definición de los indicadores de salud y comportamiento.

Actualmente, la falta de directrices sobre cómo proceder al combinar la información es particularmente problemática, ya que gran parte del proceso de psicodiagnóstico se basa en el uso de instrumentos de evaluación estructurados y semiestructurados que se caracterizan por generar matrices de datos con muchos campos vacíos. Las entrevistas estructuradas implican un proceso de evaluación estandarizado en el cual se especifican claramente tanto el proceso y la estructura de la evaluación como la información que debe registrarse y la forma de hacerlo. Pero cuando estas pruebas se corrigen mediante ordenador, es habitual que los programas diseñados a tal efecto sean extremadamente rígidos en la forma en que combinan la información, o bien que no sean sensibles a la presencia de campos vacíos. Este es el caso de la entrevista DICA-III-R informatizada en versión inglesa, una entrevista muy utilizada durante los últimos años para estudios psicodiagnósticos en población infantil y juvenil. Para comprobar la ineficacia de las definiciones de sintomatología y diagnóstico de dicha entrevista, el usuario sólo tiene que grabar un caso hipotético dejando en blanco todas las preguntas referidas a un trastorno concreto. Posteriormente, dicho usuario seguramente se sorprenderá al solicitar los resultados de la evaluación diagnóstica, ya que el programa informa de que todos los síntomas y el diagnóstico del trastorno están ausentes. En el capítulo 5 de este trabajo presentamos algoritmos corregidos que automatizan la construcción de los diagnósticos contemplados el DSM-IV.

Finalmente, cuando se definen las variables que indican la presencia y/o el grado de psicopatología es importante efectuar otra importante diferenciación que hasta ahora no habíamos contemplado. Todo investigador encargado de la gestión de datos debe tener presente que un campo vacío no necesariamente implica que la información sea desconocida, especialmente cuando se utilizan instrumentos de evaluación estructurados. En este sentido, ya desde hace algunos años, diversos autores destacan la importancia de diferenciar entre los tipos de "vacío" que pueden aparecer en un campo de registro. Por ejemplo, Sonquist y Dunkelberg (1977) ya indicaban la conveniencia de distinguir entre valor ausente (desconocido) y valor no aplicable. El primero (valor desconocido) designaría aquella información requerida para la determinación del estatus diagnóstico de los sujetos, pero que es desconocida en el momento de efectuar la evaluación. De otro lado, el valor no aplicable designaría aquella información que no es pertinente obtener de un sujeto (por ejemplo no tiene sentido preguntar a un informante del sexo masculino por el inicio de la menarquía). Lamentablemente, casi todos los sistemas gestores de bases de datos y los sistemas de análisis estadístico asignan el mismo código al valor desconocido y al no aplicable (generalmente el código de "vacío"), a pesar de que las implicaciones prácticas y el tratamiento posterior del que son objeto son absolutamente diferentes. En este sentido, por ejemplo, la frecuencia de valores ausentes determina en gran medida la validez externa del estudio, en tanto que la frecuencia de valores no aplicables sólo es un índice del grado de estructuración del protocolo de evaluación. Considerar estas diferencias también es fundamental en la fase de combinación y manipulación de datos, especialmente en la fase de definición de síntomas a partir de instrumentos de evaluación con formato estructurado.

### **3.3.7. Mejora de la calidad durante el análisis estadístico**

Aunque habitualmente el análisis estadístico se considera un aspecto hasta cierto punto desvinculado de la fase del diseño de la base de datos, en este trabajo consideramos importante destacar cuáles son los factores básicos, relacionados con dicho análisis, que pueden atentar contra la calidad de los datos de una investigación. Creemos justificada esta

actitud al considerar que el análisis estadístico no deja de formar parte integral del proceso de gestión de los datos de un estudio.

En la actualidad se han definido diferentes modalidades del efecto GIGO, aunque no todas son igualmente reconocidas como tales. Una de las que tal vez ha recibido menos atención se refiere a la falta de calidad de los programas o instrucciones que se utilizan para efectuar los análisis estadísticos de los datos de un estudio (Cobos, 1995). En este sentido, los problemas fundamentales acontecen cuando los usuarios de dichos programas o técnicas de análisis poseen conocimientos superficiales, tanto en dichos sistemas como en metodología, estadística y proceso de datos.

Curiosamente, los diseñadores de "software" se han preocupado de hacer los programas habituales accesibles a una gran cantidad de usuarios, de tal manera que en la actualidad es relativamente sencillo, por ejemplo, ajustar modelos de regresión logística sin conocer qué forma sigue la función logística (Cobos, 1995). Por nuestra parte consideramos que este tipo de deficiencia es, si es que se nos permite la expresión, "menos trascendente" que los problemas de calidad en el proceso de manejo y combinación de los datos. No pretendemos crear la idea de que la técnica de análisis escogida sea irrelevante o la interpretación que de ella se realice; bien al contrario, ambos aspectos constituyen elementos fundamentales de cara a la validez de las conclusiones de cualquier estudio. Sucede, sin embargo, que es un problema relativamente más fácil de solucionar una vez se ha detectado. En efecto, cuando se manejan "datos de calidad", el análisis estadístico, por vasto y complejo que parezca, se acostumbra a resolver sin mayores dificultades buscando el consejo de expertos. El problema fundamental acontece cuando se manejan datos que no han sido sometidos a ningún control rutinario de calidad. En estos casos, ni el mejor programa de análisis de datos ni el mejor estadístico sirven de gran ayuda.

Finalmente, nos gustaría destacar que el análisis estadístico ofrece otra interesante oportunidad para detectar valores anómalos, fuera de rango o inconsistentes no detectados en fases anteriores del proceso de datos. En este sentido, algunos autores también proponen realizar controles de consistencia en relación a los análisis estadísticos que se efectúan, con el objeto de detectar y corregir los posibles errores todavía pudiera contener la información que se maneja (Gassman et al., 1994).

### **3.3.8. Cómo se puede evaluar el proceso de calidad de la gestión de datos clínicos**

Haber adoptado los controles que se presentan en los apartados anteriores no garantiza totalmente la calidad de la información y de la gestión de datos de una investigación. Por esta razón, en la literatura se encuentran referenciados procedimientos y controles especialmente concebidos para evaluar la calidad de la gestión como producto (Connett y Lee, 1990), volviendo a la idea original de que nunca se deberían analizar datos de los que no se dispone de una estimación del error que contienen. En los siguientes apartados abordamos el problema de la evaluación del proceso de calidad del manejo de datos. Para ello vamos a considerar que la implementación de cualquier proceso de calidad se efectúa de acuerdo a cuatro etapas fundamentales (DISA, 1997): (1) la identificación de los requisitos para garantizar la calidad de los datos y la definición de los problemas relacionados con la calidad de la gestión; (2) la medición de la calidad de la gestión; (3) el análisis "cualitativo" de los problemas evidenciados en la fase anterior; y (4) la valoración de las estrategias oportunas para incrementar la calidad de la gestión en futuros trabajos.

### ***3.3.8.1. Requisitos de calidad y definición de problemas***

Definir tanto los requisitos de calidad como los problemas específicos que pueden estar relacionados con la gestión de los datos de una investigación no es, en absoluto, trivial. Esta tarea requiere que el investigador conozca todas aquellas dificultades que pueden haber surgido durante el proceso de gestión de los datos que maneja, independientemente de que se hayan o no introducido los correspondientes controles de calidad. De este modo, si durante las fases anteriores del proceso de datos se han incorporado dichos controles, es importante contrastar en qué grado han sido efectivos. Y si en las fases previas del proceso de datos no se ha introducido ningún control es fundamental evaluar el grado de consistencia de la información que se maneja durante el análisis estadístico y el establecimiento de las correspondientes conclusiones.

En cualquier caso, puesto que estos aspectos suelen ser poco conocidos, tanto la especificación de los requisitos de calidad como la definición exhaustiva de los posibles problemas acaecidos implica que los investigadores recurran, necesariamente, a fuentes documentales o bibliográficas especializadas en calidad de la gestión.

Y en relación a este punto también es importante tener presente que tanto la definición de los requisitos de calidad como de los posibles problemas de gestión también conlleva asociada la revisión del programa informático que se ha utilizado para manejar los datos. Es importante que el investigador tenga presente en todo momento tanto las prestaciones del software para optimizar la calidad del proceso como las deficiencias del sistema en lo que se refiere a manejo o gestión de la información. Ambos aspectos inciden de forma determinante en la valoración de la calidad de los datos de un estudio.

Finalmente, simplemente recordar que los problemas que pueden aparecer durante la gestión de datos son de naturaleza muy diversa. La revisión de las diferentes fuentes bibliográficas señalan que gran parte de los errores son debidos a deficiencias en el registro de la información y en el posterior manejo de los datos (DISA, 1997). Sin embargo, tal como hemos indicado en el párrafo anterior, no es posible descartar que parte de las deficiencias sean debidas al sistema informático del que se dispone. En este sentido, por ejemplo, los usuarios podrían estar utilizando los sistemas de gestión que existen comercialmente de forma inadecuada (por ejemplo debido a la carencia de la documentación necesaria, o simplemente a la falta de entrenamiento en dichos sistemas), o incluso podrían estar utilizando programas que no han sido probados adecuadamente en fases previas por los correspondientes diseñadores de "software". En relación a este último punto, la normativa del Comité Internacional de Editores de Revistas Biomédicas, recogida en los "Requisitos de Uniformidad para manuscritos presentados a revistas biomédicas", señala la conveniencia de especificar el software y la versión utilizada para el análisis de datos. De la misma forma, la fase de diseño de la base de datos podría ser inadecuada a los propósitos de un estudio, o contener inconsistencias que favorecieran posteriores errores de manejo de datos.

### ***3.3.8.2. Medición de la calidad de la gestión***

Una vez concretados los problemas que pueden atentar contra la calidad del proceso de gestión de datos en una investigación, es importante evaluar la magnitud en que dichas dificultades están o no presentes. En este apartado recogemos los controles básicos utilizados en la práctica al hablar de la calidad de la gestión de datos como un producto final y de los estimadores que evalúan en qué grado los datos consiguen dicha calidad. En este caso es habitual diferenciar dos conceptos: la verificación y la certificación.

En el primer caso, la validación de los datos grabados comporta, principalmente, verificar que la información contenida en la base de datos corresponde a la información registrada en los documentos originales. Para ello se implementan y utilizan procedimientos

orientados a garantizar que tanto la base de datos como la información contenida cumplen con las restricciones establecidas en las fases anteriores del diseño, y en el caso de que se detecten errores se corrijan. En este sentido, por ejemplo, se pueden obtener valoraciones sobre la eficacia y eficiencia de los operadores de entrada de datos a través de sistemas de doble entrada o de verificación aleatoria de un porcentaje total de los datos (Blumenstein, 1993; Doménech, Losilla y Portell, 1997; Gassman et al., 1995; Gibson et al., 1994; Wolf, 1993). De otro lado, algunos autores también proponen implementar procesos de valoración de la calidad del registro de datos "a posteriori", es decir después de que la información haya sido grabada y antes de proceder a la generación de las variables necesarias para efectuar los análisis estadísticos. En este caso se argumenta que a pesar de que se hayan incorporado todos los controles posibles durante la grabación y el manejo de la información, siempre es adecuado disponer de un programa para chequear que los datos que serán objeto de análisis y/o interpretación no contienen todavía valores incorrectos (Fowler, 1993). Estos chequeos "a posteriori" permiten detectar los errores de registro y gestión que podrían haber pasado desapercibidos durante las fases anteriores, y facilita la obtención de otros indicadores de calidad de los datos que serán objeto de análisis e interpretación. Asimismo, se considera que los chequeos de consistencia "a posteriori" son absolutamente indispensables cuando la grabación de la información se ha efectuado a través de programas que no han sido diseñados para garantizar la calidad del registro, y que por lo tanto carecen de los controles y protecciones propios de los gestores especializados en bases de datos. En este último caso, puesto que los sistemas informáticos no disponen de procedimientos automatizados para efectuar la "depuración" de datos y la valoración de la calidad del registro, se han realizado propuestas procedimentales para ayudar a los usuarios de dichos sistemas a detectar los valores inconsistentes, cambiar los datos erróneos por valores correctos (si es que se dispone de fuentes con la información correcta) y obtener las correspondientes estadísticas de incidencias (Doménech, 1998).

En relación a la valoración de la calidad de la gestión, además, de los indicadores de la magnitud de los problemas que se relacionan con la gestión de datos es imprescindible disponer de información sobre el tipo de error cometido (a un nivel más cualitativo) y de las causas más probables que pueden estar favoreciendo dichas deficiencias. En el primer caso, el estudio del tipo de error cometido (por ejemplo durante el registro de los datos) implica responder dos preguntas fundamentales: (1) en qué áreas se acumula un mayor porcentaje de error; y (2) si ciertos tipos de error ocurren con mayor frecuencia. Disponer de esta información es fundamental, ya que orienta a los investigadores sobre las áreas en que es importante concentrar mayores esfuerzos para asegurar mínimamente la calidad de la gestión de datos en futuros trabajos. Disponer, además, de las causas explicativas de la distribución y del tipo de error, les orientará sobre las estrategias particulares y los controles que se deberían adoptar para reducir dichas incidencias.

Finalmente, la certificación constituye el conjunto de estrategias que aplica el organismo responsable de llevar a cabo la investigación para establecer en qué grado el diseño de la base de datos y/o la información contenida cumplen con los objetivos planteados en dicha investigación. En este trabajo no abordamos este aspecto de calidad debido a que las acciones que comporta dependen de los objetivos particulares de cada investigación.

### ***3.3.8.3. Valoración de los costes de adoptar estrategias para mejorar la calidad de la gestión***

El análisis de la calidad de los datos no está completo hasta que se valoran las estrategias que se podrían adoptar para mejorar la calidad de la gestión en futuros trabajos de investigación (Gassman et al., 1995). En este sentido, es importante explicitar y evaluar tanto

los costes de corregir (o no) las fuentes de error como la adopción de soluciones para mejorar el proceso de gestión adoptado.

En el primer caso, la valoración de los costes que conlleva corregir o no las fuentes que contribuyen a los errores de gestión no es una tarea fácil. En general, dichos costes se suelen categorizar en diferentes áreas (DISA, 1997):

COSTES DIRECTOS	COSTES INDIRECTOS
1. Costes controlables 1.1 Costes de prevención de error 1.2 Costes de valoración de error 1.3 Costes de corrección de error 2. Costes de equipamiento y entrenamiento 3. Costes resultantes de la baja calidad de la gestión de datos	1. Costes relacionados con la insatisfacción del investigador 2. Costes relacionados con la falta de credibilidad de los resultados del estudio

Los errores directos clasificados como controlables constituyen todos aquellos costes reales invertidos en la prevención, la valoración y la corrección de los errores de gestión. De la misma forma, los costes de equipamiento y entrenamiento indican los costes que supone adoptar el "hardware" y el "software" necesario, y también el adecuado entrenamiento del personal, para efectuar dichas tareas de prevención, valoración y corrección de los errores de proceso. Al evaluar este tipo de costes es habitual comparar dos o más alternativas ideadas para mejorar la calidad de datos. La valoración de dichas acciones correctivas se efectúa de acuerdo a los recursos reales (fundamentalmente en lo que se refiere a aspectos económicos, de personal investigador y de tiempo disponible) con los que cuenta cada proyecto o estudio. Asimismo, cuando existen varias alternativas mutuamente excluyentes, es importante valorar conjuntamente los beneficios y el riesgo que conlleva la adopción de cada una.

Los costes resultantes de la baja calidad de la gestión y los costes indirectos suelen ser más difíciles de cuantificar que los anteriores. Sin embargo, siempre deberían ser considerados a la hora de evaluar el impacto de la baja calidad de la gestión de datos, ya que pueden ser en sí mismos determinantes para incorporar modificaciones sustantivas en todo el proceso de gestión de datos y en las estrategias utilizadas para garantizar la calidad.

Finalmente, en lo que se refiere a la propuesta de soluciones para mejorar la calidad de los datos, reiteramos que es sumamente interesante que los grupos que trabajan dentro de una determinada línea de investigación contemplen la posibilidad de adoptar, junto al proceso de gestión de datos, procesos de calidad de la gestión. Dichos procesos de calidad, aunque necesariamente dependerán de las características del entorno de trabajo específico donde sean adoptados, seguirán unas fases secuenciales claramente delimitadas (DISA, 1997): (1) establecer las características del entorno que determinarán los requisitos del proyecto de calidad de gestión; (2) desarrollar dicho proyecto, a nivel de objetivos y estrategias para garantizarlos; (3) implementar el proyecto; y (4) evaluar en qué grado el proceso adoptado satisface los requisitos planteados a nivel de eficacia y eficiencia, así como también qué aspectos de calidad de gestión no han sido adecuadamente satisfechos.

### 3.3.9. Síntesis de las estrategias de control propuestas

Queremos terminar este apartado sintetizando las estrategias básicas de control que hemos propuesto en esta parte del capítulo con objeto de mejorar la calidad de la gestión de datos clínicos. Para ello hemos construido una figura (figura 3.3) donde se recogen las fases básicas que conlleva el proceso de datos (tomando como referencia el enfoque del diseño de bases de datos) junto a las estrategias elementales de control de calidad.

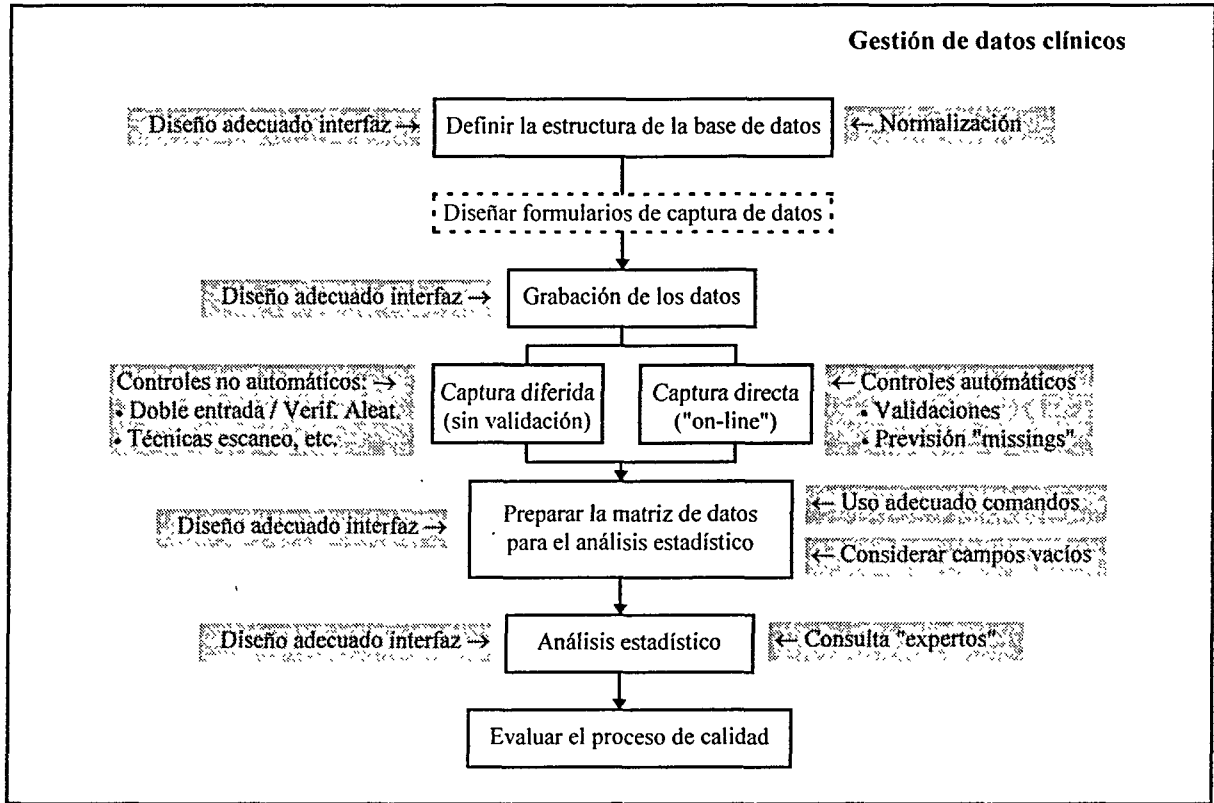


Figura 3.1. Controles básicos de calidad en la gestión de datos clínicos.

# UN ESTUDIO DE LA CALIDAD DE LA CAPTURA DE DATOS PSICODIAGNÓSTICOS

*"Quality control of coding includes check-code a sample of each coder's work. This serves two purposes: it identifies coders who are making coding decision errors and it identifies coding rules that are ambiguous and require clarification".*

(Floyd J. Fowler, 1993)

## 4.1. INTRODUCCIÓN: JUSTIFICACIÓN DE LA EXPERIENCIA Y OBJETIVOS

El proceso de captura de la información en formato informatizado constituye una de las etapas fundamentales para la construcción de la matriz de datos sobre la que se realizarán los correspondientes análisis estadísticos. Los datos aportados en el capítulo previo indican que la investigación actual atribuye una gran relevancia al estudio de las fuentes de error que afectan la calidad del registro, y también a las estrategias de control que se pueden adoptar para mejorarla.

Nuestro equipo de investigación viene trabajando durante los últimos años en el ámbito del diagnóstico psicológico infantil y juvenil, y la necesidad de transcribir al ordenador los datos obtenidos a través de entrevistas estructuradas y cuestionarios autoadministrados (ambos en formato de papel), ha constituido durante mucho tiempo nuestra realidad diaria. Esta situación, común a la de muchos otros equipos de investigación en nuestro país, ha suscitado en nosotros un creciente interés por todos aquellos aspectos que se relacionan con la calidad del registro, interés que se ha concretado en un estudio empírico para valorar la calidad de la captura de datos en el ámbito del psicodiagnóstico. Por la información de que disponemos hasta el presente, en muchos aspectos este estudio carece de antecedentes en nuestro país.

Con el propósito de alcanzar los objetivos propuestos, hemos diseñado una experiencia que consiste en transcribir al ordenador, en ausencia de controles, un conjunto de entrevistas estructuradas que previamente habían sido codificadas en papel por diferentes entrevistadores. En este capítulo presentamos dicho estudio junto a las principales conclusiones que se derivan. Consideramos que la experiencia propuesta reproduce en gran medida la estrategia procedimental habitual que subyace a un gran número de trabajos que se llevan a cabo en nuestro país. Por consiguiente, las medidas obtenidas en relación a los errores de grabación que comete un operador profesional en el registro de datos, unido al estudio de la distribución y magnitud de dichos errores, nos permitirán obtener indicadores de calidad de los datos aportados en dichos estudios de investigación. La experiencia también nos permitirá determinar qué variables pueden estar relacionadas con dichos aspectos de calidad.

En síntesis, este capítulo aborda la problemática de la valoración de los errores que se producen cuando se capturan datos en ausencia de protecciones automáticas. En estos casos es imprescindible disponer de un indicativo de la calidad de los datos, prácticamente imposible de obtener cuando se realiza una simple entrada. Tal como se indicó en el capítulo anterior, la estrategia habitual frente a este problema suele comportar la doble entrada de los datos, a pesar de que este procedimiento raramente es utilizado debido al incremento que supone sobre el coste de la investigación. En este trabajo se valora cómo funciona una alternativa que consiste en la verificación aleatoria de un porcentaje del total de los registros, escogidos al



azar por el sistema informático y presentados con un retardo suficiente para evitar efectos de recuerdo, frente a la doble entrada (con y sin "feedback"). En los siguientes subapartados se explicitan los objetivos específicos de la experiencia.

#### **4.1.1. Primer objetivo de la experiencia**

El objetivo básico de esta experiencia ha consistido en comparar las dos estrategias básicas de control de la calidad del registro de datos: la "verificación aleatoria" frente a la doble entrada (con y sin "feedback"). Para ello se analizarán las discrepancias que comete una operadora profesional durante el registro de datos, a un nivel tanto cuantitativo como cualitativo. En este segundo caso, se presta especial atención a la distribución de las discordancias en función del tipo de campo, de la sección de la entrevista, del tipo de protocolo, y del entrevistador que originalmente había codificado los datos en papel.

Consideramos que en la medida en que se alcance este primer objetivo, se obtendrá información relevante sobre dos aspectos básicos:

1. Qué estrategia de control de calidad del registro obtiene menos discrepancias.
2. Detectar si existe algún elemento que favorezca la acumulación de un mayor porcentaje de discrepancias (tipo de campo, sección de la entrevista, tipo de protocolo o entrevistador original). En el caso de que existieran diferencias, se obtendrían indicios sobre los factores que actúan dificultando el registro a los operadores encargados de la captura de datos (y/o también a los entrevistadores que previamente habían codificado la información en papel).

#### **4.1.2. Segundo objetivo de la experiencia**

El segundo objetivo básico de esta experiencia ha consistido en valorar qué errores persisten cuando se utilizan la verificación aleatoria y la doble entrada como técnicas de control de la captura de datos. Para ello se comparan la entrada de datos finales verificados con una entrada o patrón de referencia (ausente de errores). Las diferencias entre ambas entradas, a nivel cualitativo y cuantitativo, aportarán evidencias sobre:

1. El tipo de estrategia que resulta más idónea para la captura de datos, en la medida en que obtiene menor porcentaje de errores.
2. Detectar, si es que existen, qué elementos favorecen la acumulación de un mayor porcentaje de errores.
3. El tipo de errores que persisten tras la verificación de los datos. Básicamente se compararán dos tipos de error: la omisión de valores frente al cambio de códigos.

## **4.2. MÉTODO**

### **4.2.1. Sujetos**

Las condiciones de inclusión de los operadores de entrada de datos en el estudio consistían en:

1. Un mínimo de 1 año de experiencia en la grabación de datos clínicos.
2. Demostrar eficiencia en la grabación de datos. En particular, se exigió un ritmo de grabación superior a 30 palabras por minuto sin error.
3. Disponer del tiempo que requería el registro verificado de las entrevistas.

Para la realización de esta experiencia, debido a las características del procedimiento utilizado, hemos procedido a trabajar con un número reducido de operadores de entrada de datos. En particular, contamos con la colaboración de dos operadores de entrada de datos:

1. Una operadora principal, con varios años de experiencia en el proceso de transcripción de datos clínicos al ordenador, y con un ritmo promedio de grabación igual a 40 palabras por minuto sin error.
2. Asimismo, y puesto que diferentes trabajos sugieren que la doble entrada debe realizarse por un operador diferente al que llevó a cabo la primera entrada, en esta experiencia también hemos contado con un segundo operador encargado de realizar otra doble entrada de los cuestionarios. Se aseguró que este segundo operador también cumplía las condiciones de inclusión en el estudio.

### **4.2.2. Material**

Para realizar esta experiencia se disponía de 380 entrevistas estructuradas que previamente habían sido codificadas en papel, y cuyos datos debían ser capturados en formato informatizado. La información contenida en dichas entrevistas correspondía a una muestra clínica formada por familias que acudieron (por voluntad propia) a la consulta de distintos Servicios de Psicología y Psiquiatría Infantil de hospitales de la red pública barcelonesa, así como también por sujetos escolarizados de la población general. Estos datos habían sido obtenidos entre octubre de 1993 y abril de 1996.

En concreto, estos datos habían sido recogidos mediante la entrevista *Diagnostic Interview for Children and Adolescents-Revised* (DICA-R, Reich et al., 1991). La DICA, cuya estructura se presenta ampliamente en el siguiente capítulo, había sido utilizada para comprobar la existencia de un diagnóstico psicopatológico en los sujetos, obteniendo los informes de los padres (o en su defecto, de los principales cuidadores) y del propio sujeto (niño/adolescente). En los Anexos 1A y 1B se presentan dos ejemplos con las hojas de registro utilizadas en una de las familias del estudio, en concreto para la entrevista de padres y para la de hijos. Es importante destacar que si los informantes eran los padres (o principales cuidadores), se recababa información sobre todos los hermanos con edades comprendidas entre 6 y 17 años (edades de aplicación de la entrevista DICA-R) que convivían en la misma casa.

Para realizar esta experiencia, hemos preparado un formulario informatizado con el Sistema DAT (Doménech y Losilla, 1995). El formulario elaborado, que constaba de 983 campos, carecía de protecciones para la captura de datos (no se definieron validaciones ni saltos), de modo que la operadora podía introducir cualquier tipo de valor en cualquier campo siempre y cuando éste fuera consistente con el formato definido. Asimismo, y para facilitar el registro de la información, diseñamos las pantallas de forma que la disposición de los campos de

captura reprodujeran lo más fielmente posible el aspecto del protocolo en papel que contenía los datos originales.

### 4.2.3. Diseño

El experimento se ha realizado simulando el entorno de trabajo de un equipo de investigación. En concreto, las sesiones de entrada de datos se han realizado en el Laboratori d'Estadística Aplicada i de Modelització de la Unviersitat Autònoma de Barcelona, utilizando un ordenador PC-Pentium2, con monitor color de 14" y tarjeta gráfica VGA.

La variable independiente del estudio ha consistido en el tipo de registro para los datos, tal como ilustra la figura 4.1.

La condición ES se ha caracterizado por una consigna verbal informando de la importancia de llevar a cabo una captura correcta de los datos que estaban registrados en los formularios de papel.

La condición VA se ha caracterizado por una consigna verbal informando a la operadora que el programa controlaría los errores cometidos durante la entrada de datos, obligándola a repetir la introducción de algunos registros, indicándole si se cometían discrepancias respecto a la primera entrada y solicitándole que verificase los campos con discordancia. El porcentaje de formularios a verificar se ha fijado en un 20%, con un ratardo en la doble entrada igual a 2 formularios.

La codición DE (con "feedback") se ha caracterizado por una consigna verbal informando de que el programa volvería a solicitar la introducción de todos los registros introducidos al inicio, de forma similar a la verificación aleatoria.

La condición DE (sin "feedback") se ha caracterizado por la misma consigna verbal que la ES.

Se han definido dos variables dependientes en función de los objetivos de la experiencia: el total de discrepancias que se producen entre las Fases A y B, y el total de errores que se producen al comparar los datos verificados con la entrada de referencia (patrón).

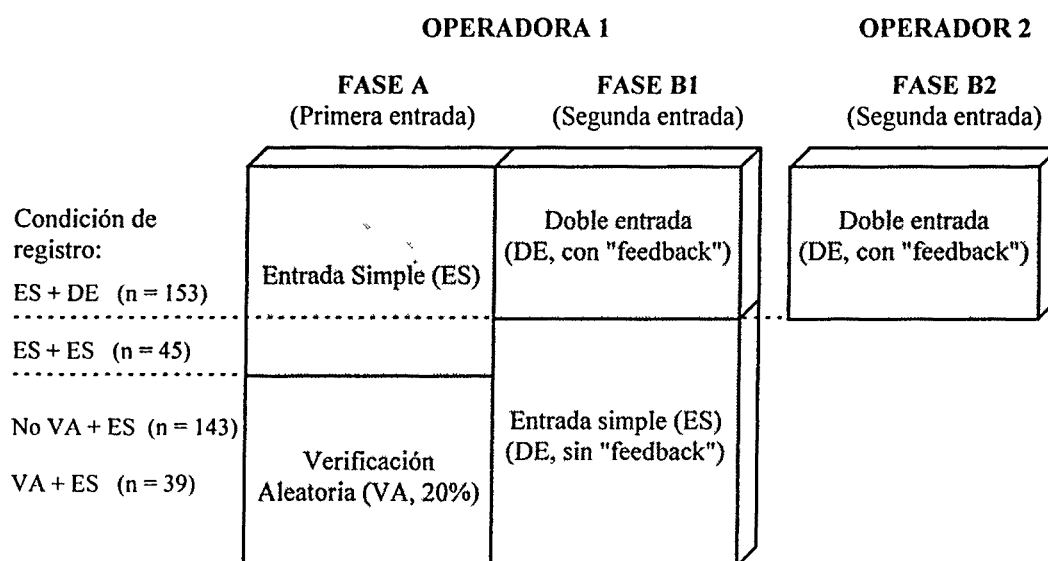


Figura 4 1 Esquema del diseño experimental para el estudio

#### 4.2.4. Procedimiento

Esta tarea no ha sido presentada como un experimento, sino como una colaboración en un trabajo de investigación real llevado a cabo por el equipo de investigación de la Universitat Autònoma de Barcelona.

La experiencia se ha estructurado en dos etapas bien diferenciadas:

1. Entrenamiento: informando a los operadores de los objetivos del registro de formularios. A continuación se les presentaba el programa informatizado y se les ayudaba a que practicasen con él la entrada de algunos cuestionarios de prueba, y a que realizaran las copias de seguridad una vez finalizaran la entrada de la información. De este modo, los operadores tomaban contacto con el software de entrada de datos y se familiarizaban con los protocolos. En la fase de entrenamiento también se informó a los operadores de que podían realizar los descansos que necesitaran durante cada sesión, y se incentivó la rapidez de la entrada de datos.
2. Fase experimental: finalizada la fase de entrenamiento, se iniciaron las sesiones de captura de datos. A medida que avanzaba el registro, el responsable de llevar a cabo la fase experimental ofrecía las correspondientes consignas verbales (ES, VA y DE). Cada sesión de captura de datos duraba en promedio, de 6 a 8 horas. Cada operador podía realizar descansos regulares de acuerdo a sus necesidades, según su propia voluntad. Durante el tiempo de entrada de datos, no se realizó ninguna otra actividad en el Laboratori d'Estadística. Asimismo, la operadora que ha llevado a cabo la entrada de las Fases A y B1 (figura 4.1) no fue informada de que los datos volverían a ser entrados por otro operador (Fase B2, figura 4.1).

Al final de la fase experimental, se disponía de las copias realizadas por el Sistema DAT de las correspondientes etapas en las que se estructuró la experiencia (tabla 4.1):

*Tabla 4.1. Esquema de las copias obtenidas del Sistema DAT para el análisis de los datos.*

Copia	Fase	Condición de registro
1	A	Entrada simple (n = 153 + 45)
2	A	Verificación aleatoria (n = 143 + 39)
3	B	Doble entrada (n = 153)
4	B1	Entrada simple (n = 45 + 143 + 39)
5	B2	Doble entrada (n = 153)

#### 4.2.5. Análisis de los datos

El análisis estadístico de los datos se ha efectuado con el Sistema SPSS 7.5.2 para Windows. Previamente, se ha procedido a elaborar la matriz objeto de análisis de acuerdo a los pasos que exponemos en los siguientes párrafos. Debido a la complejidad de las transformaciones requeridas y al volumen de datos que participaba en dichas transformaciones, gran parte del análisis de datos se ha realizado a través de macros que automatizan dicho proceso y garantizan la ausencia de errores. En el Anexo 2 reproducimos los procedimientos básicos que hemos utilizado en este análisis.

Los datos disponibles al finalizar este estudio, cuya estructura se ha resumido en la figura 4.1, nos han permitido plantear las comparaciones que aparecen destacadas en la parte superior de la figura 4.2. Estas comparaciones permiten abordar los objetivos planteados al diseñar la experiencia.

Para poder realizar estas comparaciones, se ha elaborado una matriz de datos, similar a la que hemos dibujado en la parte central de la figura 4.2, donde constan las variables identificadoras de cada entrevista entrada junto a las variables que miden las coincidencias/discrepancias y los errores/aciertos. Las variables identificadoras han sido:

1. Identificador del registro: con 4 valores, según la condición de registro (ES+DE, ES+ES, NVA+ES y VA+ES).
2. Tipo de protocolo: con dos valores, padres frente a hijos.
3. Entrevistador que codificó previamente la entrevista en papel.

Las variables que miden las coincidencias/discrepancias entre las Fases A y B se han obtenido de diferente forma:

- a. En el caso de los registros cuya verificación era gestionada por el Sistema DAT, es decir las entrevistas con verificación aleatoria (n=39) y doble entrada (n=153), el gestor nos ha proporcionado directamente una nueva matriz donde cada una de las variables de la entrevista DICA-R aparece codificada con dos valores según haya habido coincidencia o discrepancia al volver a entrar los datos en los correspondientes campos. Esta matriz permite obtener las variables que miden el total de coincidencias/discrepancias directamente a partir de recuentos. Sin embargo, debido a que la DICA posee una estructura ramificada que hace que sólo se complete una parte de la información para cada entrevista, hemos creído conveniente obtener las variables que miden el total de coincidencias/discrepancias en función de cómo aparecen los campos en el patrón de referencia. En este sentido, se han obtenido las coincidencias/discrepancias según el patrón estuviera completo (con valor) o vacío.

El procedimiento que hemos seguido para obtener las variables intermedias queda ilustrado en la figura 4.3. Se trata de generar para cada variable de la DICA-R una variable intermedia con 4 valores: hay coincidencia y el patrón está completo, hay coincidencia y el patrón está vacío (missing), hay discrepancia y el patrón está completo, y hay discrepancia y el patrón está vacío. Para ello se ha utilizado la macro que aparece reproducida en el Anexo 2.

Finalmente, las variables que recogen las coincidencias/discrepancias totales se han obtenido a través de recuentos de los códigos asignados a las variables intermedias. En concreto, se han obtenido los siguientes recuentos:

1. Para toda la entrevista.
2. Para las diferentes secciones en que se estructura la entrevista: demográficos, trastorno por déficit de atención con hiperactividad, negativismo, trastorno de conducta, alcohol, tabaco, colas, marihuana, otras drogas, depresión mayor,

episodio maniaco, ansiedad de separación, trastorno por evitación, ansiedad excesiva, fobias, trastorno obsesivo-compulsivo, trastorno por estrés postraumático, anorexia, bulimia, enuresis, encopresis, menstruación, identidad sexual, somatización, psicosis, estresores y observaciones.

3. Para los diferentes tipos de campo: formato numérico, fecha, momento temporal y multirresposta<sup>6</sup>.

b. En el caso de los registros cuya verificación no era gestionada por el Sistema DAT, es decir las entrevistas con entrada simple (n=45 y n=143), el gestor no nos ha proporcionado la matriz de coincidencias/discrepancias. En este caso, ha sido preciso efectuar la comparación manualmente a través de una macro SPSS (Anexo 2), tal como queda indicado en la figura 4.4. Las variables finales que recogen el total de coincidencias/discrepancias se han obtenido a partir de recuentos de los códigos asignados a dichas variables intermedias, tal como se ha indicado en el punto anterior (para toda la entrevista, para cada sección y para cada tipo de campo).

De otro lado, las variables que aparecen en la figura 4.2 que miden los errores entre los datos verificados (Fase B) y los patrones, se han obtenido de la siguiente forma:

- a. En el caso de los registros cuya verificación era gestionada por el Sistema DAT, es decir las entrevistas con verificación aleatoria (n=39) y doble entrada (n=153), hemos comparado la matriz con los datos finales facilitados directamente por el gestor con la matriz correspondiente al patrón de referencia. La figura 4.5 esquematiza el proceso seguido para llevar a cabo esta comparación (la macro utilizada figura en el Anexo 2). Las variables con los errores totales también se han obtenido a partir de recuentos para toda la entrevista, para cada sección y para cada tipo de campo.
- b. En el caso de los registros cuya verificación no era gestionada por el Sistema DAT, es decir las entrevistas con entrada simple (n=45 y n=143), se ha procedido a la elaboración de la matriz de datos verificados. Para ello se ha facilitado a la operadora un listado con los campos en los que se habían producido discrepancias entre las Fases A y B, y se le ha solicitado que registre el valor correcto en dichas variables. A continuación se ha procedido a la comparación de dicho archivo de datos verificados con el correspondiente patrón de referencia, tal como se ilustra en la figura 4.6. La macro utilizada para la comparación es la misma que la utilizada en la figura 4.5. Finalmente, las variables con los errores totales se han obtenido como en los casos anteriores: a partir de recuentos para toda la entrevista, para cada sección y para cada tipo de campo.

Una vez obtenida la matriz con las variables de coincidencias y error para cada registro (parte central de la figura 4.2) se ha procedido a elaborar la matriz que recoge las sumas de dichas variables para las siguientes agrupaciones:

1. Identificador del registro: con 4 valores, según la condición de registro (ES+DE, ES+ES, NVA+ES y VA+ES).
2. Tipo de protocolo: con dos valores, padres frente a hijos.
3. Entrevistador que codificó previamente la entrevista en papel.

---

<sup>6</sup> Los campos multirresposta permiten registrar en un sólo campo diferentes valores, en el caso del Sistema DAT a través de una tabla. Estos campos son adecuados para registrar diferente tipo de información repetida y cuya frecuencia es variable de un registro a otro, por ejemplo los trastornos mentales que han padecido los miembros familiares. En el siguiente capítulo se revisa ampliamente el concepto de campo multirresposta.

4. Sección de la entrevista.

5. Tipo de campo.

Para elaborar esta matriz se han utilizado las funciones "AGGREGATE" y "ADD FILES" del Sistema SPSS, tal como se reproduce en el Anexo 2. Los listados que se incluyen en el apartado de resultados han sido obtenidos a partir de las matrices finales con datos agregados, ponderando según la variable requerida en cada caso.

Registro:	Primer objetivo: Estudio de discrepancias		Segundo objetivo: Estudio de errores		
	Comparación FA <sup>1</sup> - FB1 <sup>2</sup>	Comparación FA-FB2 <sup>3</sup>	Comparación FA-PATRÓN <sup>4</sup>	Comparación FB1-PATRÓN	Comparación FB2-PATRÓN
ES + DE	✓	✓	✓	✓	✓
ES + ES	✓	-	✓	✓	-
NVA + ES	✓	-	✓	✓	-
VA + ES	✓	-	- <sup>5</sup>	✓	-

<sup>1</sup> Fase A; <sup>2</sup> Fase B1; <sup>3</sup> Fase B2; <sup>4</sup> Entrada de referencia; <sup>5</sup> Esta comparación no se ha podido realizar debido a que el Sistema DAT, tras efectuar la verificación aleatoria, únicamente mantiene almacenado el valor final que el/la operador/a ha considerado adecuado en cada campo



Variables identificadoras de la entrevista				Variables de discrepancias				Variables de error			
Caso	Identif. registro	Tipo protoc.	Entrevistador	X <sub>C1</sub>	X <sub>C2</sub>	...	X <sub>Cn</sub>	X <sub>E1</sub>	X <sub>E2</sub>	...	X <sub>En</sub>
1											
2											
3											
...											
...											
...											
380											



Variables de agrupación			Error	Sumatorios variables de discrepancias				Sumatorios variables de error			
Id. regist.	Tipo protoc.	Entrevistador		X <sub>C1</sub>	X <sub>C2</sub>	...	X <sub>Cn</sub>	X <sub>E1</sub>	X <sub>E2</sub>	...	X <sub>En</sub>
1			0								
2			0								
3			...								
...			...								
...			...								
.			1								
n			1								

Figura 4 2 Esquema del proceso de construcción de la matriz objeto de análisis estadístico

En cuanto al análisis realizado para comparar la eficacia de las diferentes estrategias de captura de datos y la distribución de las discrepancias y los errores, se ha utilizado dos diferentes estrategias (Doménech, 1998):

1. Puesto que la variables independiente (por ejemplo, la condición de registro o el tipo de protocolo) y la variable dependiente (por ejemplo, el porcentaje de discrepancias/errores en cada clase) se han considerado binarias, las comparaciones se han realizado utilizando el estadístico de contraste  $\chi^2$  de Pearson.
2. Para valorar la magnitud de las discrepancias se ha obtenido el intervalo de confianza de la diferencia de proporciones.

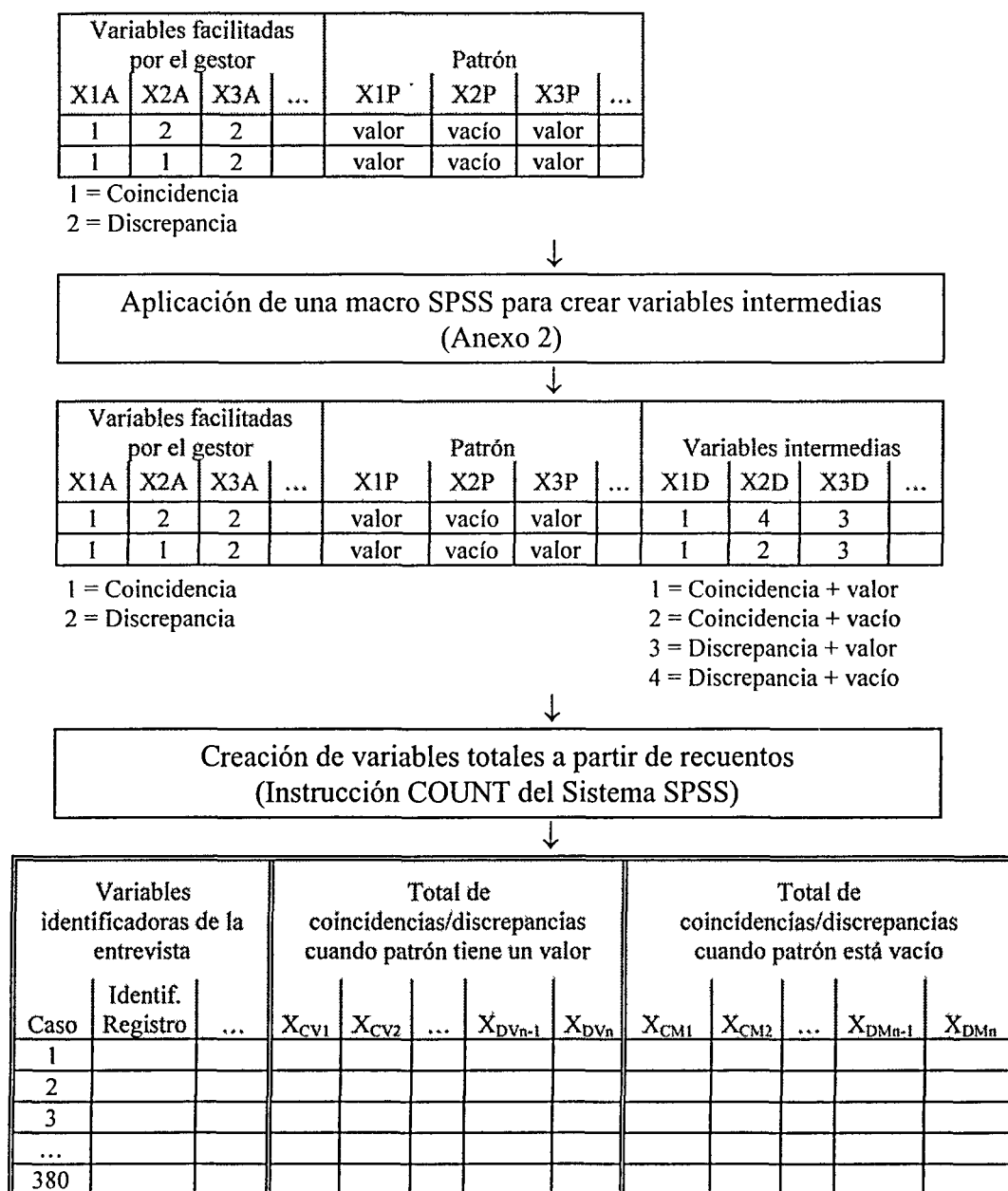


Figura 4.3 Esquema del proceso de definición de las variables que recogen las coincidencias-discrepancias totales para los registros con verificación gestionada por el Sistema DAT.



FASE A		FASE B1		PATRÓN	
X1A	...	X1B	...	X1P	...
2		2		4	
.		.		4	
2		2		.	
.		.		.	
2		1		4	
.		2		4	
2		.		4	
2		1		.	
.		2		.	
2		.		.	



Aplicación de una macro SPSS para crear variables intermedias  
(Anexo 2)



X1A	X1B	X1P	→	X1D
2	2	4		1
.	.	4		1
2	2	.		2
.	.	.		2
2	1	4		3
.	2	4		3
2	.	4		3
2	1	.		4
.	2	.		4
2	.	.		4

Las nuevas variables valen que miden discrepancias valen:

- 1 = Coincidencia + valor en patrón
- 2 = Coincidencia + vacío
- 3 = Discrepancia + valor
- 4 = Discrepancia + vacío



Creación de variables totales a partir de recuentos  
(Instrucción COUNT del Sistema SPSS)



Variables identificadoras de la entrevista			Total de coincidencias/discrepancias cuando patrón tiene un valor					Total de coincidencias/discrepancias cuando patrón está vacío				
Caso	Identif. Registro	...	X <sub>CV1</sub>	X <sub>CV2</sub>	...	X <sub>DVn-1</sub>	X <sub>DVn</sub>	X <sub>CM1</sub>	X <sub>CM2</sub>	...	X <sub>DMn-1</sub>	X <sub>DMn</sub>
1												
2												
3												
...												
380												

Figura 4.4. Esquema del proceso de definición de las variables que recogen las coincidencias/discrepancias totales para los registros con verificación no gestionada por el Sistema DAT (entrada simple).

Datos verificados		PATRÓN	
X1B <sub>v</sub>	...	X1P	...
2		4	
.		4	
2		.	
.		.	
1		4	
2		4	
.		4	
1		.	
2		.	
.		.	



Aplicación de una macro SPSS para crear variables intermedias  
(Anexo 2)



X1B <sub>v</sub>	X1P	→	X1E
2	2		1
.	.		2
2	1		3
.	2		4
3	.		5

Las nuevas variables que miden errores valen:  
 1 = Acierto + valor en patrón  
 2 = Acierto + vacío en patrón  
 3,4 = Error + valor en patrón  
 5 = Error + vacío en patrón



Creación de variables totales a partir de recuentos  
(Instrucción COUNT del Sistema SPSS)



Variables ident. de la entrevista			Total de aciertos/errores cuando patrón tiene un valor				Total aciertos/errores cuando patrón está vacío					
Caso	Identif. Registro	...	X <sub>AV1</sub>	X <sub>AV2</sub>	...	X <sub>EVn-1</sub>	X <sub>EVn</sub>	X <sub>AM1</sub>	X <sub>AM2</sub>	...	X <sub>EMn-1</sub>	X <sub>EMn</sub>
1												
2												
3												
...												
380												

Figura 4.5. Esquema del proceso de definición de las variables que recogen los errores totales para los registros con verificación gestionada por el Sistema DAT.

FASE A		...	FASE B1	
X1A			X1B	
2			2	
1			.	
2			.	
.			.	
1			1	
2			2	
2			.	
1			.	
2			2	
.			.	

→

Datos verificados		...	PATRÓN	
X1B <sub>v</sub>			X1P	...
2			4	
.			4	
2			.	
.			.	
1			4	
2			4	
.			4	
1			.	
2			.	
.			.	

↓

Aplicación de una macro SPSS para crear variables intermedias  
(Anexo 2)

↓

X1B <sub>v</sub>	X1P	→	X1E
2	2		1
.	.		2
2	1		3
.	2		4
3	.		5

Las nuevas variables que miden errores valen:

- 1 = Acierto + valor en patrón
- 2 = Acierto + vacío en patrón
- 3,4 = Error + valor en patrón
- 5 = Error + vacío en patrón

↓

Creación de variables totales a partir de recuentos  
(Instrucción COUNT del Sistema SPSS)

↓

Variables ident. de la entrevista			Total de aciertos/errores cuando patrón tiene un valor				Total aciertos/errores cuando patrón está vacío					
Caso	Identif. Registro	...	X <sub>AV1</sub>	X <sub>AV2</sub>	...	X <sub>EVn-1</sub>	X <sub>EVn</sub>	X <sub>AM1</sub>	X <sub>AM2</sub>	...	X <sub>EMn-1</sub>	X <sub>EMn</sub>
1												
2												
3												
...												
380												

Figura 4.6. Esquema del proceso de definición de las variables que recogen los errores totales para los registros con entrada simple.

### 4.3. RESULTADOS

En los siguientes apartados se presentan los resultados principales que se obtienen del estudio que hemos llevado a cabo. La exposición se estructura en diferentes partes según los objetivos planteados al inicio de este capítulo.

#### 4.3.1. Estudio de las discrepancias entre las Fases A y B1

##### Discrepancias totales en función de la condición de registro:

La siguiente tabla recoge la proporción total de discrepancias (en tantos por mil) para cada condición de registro:

Condición registro	Si patrón está vacío	Si patrón tiene un valor	TOTAL
DE (misma operadora)	4.0775	48.7464	17.0979
ES + ES (DE sin "feedback")	5.2181	29.3516	11.8910
NO VA + ES	2.3757	28.8156	8.8640
VA	1.8921	38.6562	10.8568
DE (operadores distintos)	7.1256	70.9975	25.7433

Estos resultados sugieren que, independientemente del contenido del patrón, la condición DE con operador distinto obtiene una mayor proporción de discrepancias. De el caso de que los datos sean grabados por la misma operadora en ambas ocasiones, la DE (con y sin "feedback") obtiene más discrepancias que la verificación aleatoria.

##### Distribución de las discrepancias en función del tipo de protocolo:

Comparación de la Fase A y la Fase B1: (datos en tantos por mil)

Tipo protocolo	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Hijos	2.8110	-	41.6911	-
Padres	4.8322	1.25 a 2.79	54.7603	8.92 a 17.2

Tipo protocolo	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Hijos	1.9299	-	26.7319	-
Padres	7.0371	3.71 a 6.50	31.3629	N.S.

Tipo protocolo	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Hijos	1.6808	-	15.7362	-
Padres	2.7555	0.50 a 1.65	38.4325	19.3 a 26.0

Tipo protocolo	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Hijos	1.8235	–	38.0744	–
Padres	1.9632	N.S.	39.3013	N.S.

Comparación de la Fase A y la Fase B2: (en tantos por mil)

Tipo protocolo	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Hijos	6.7032	–	58.9646	–
Padres	7.3772	N.S.	81.2543	17.4 a 27.2

Los resultados obtenidos en relación a la distribución de las discrepancias según el tipo de protocolo indican que en todas las condiciones de registro los protocolos de padres obtienen mayor proporción de discrepancias que los protocolos de hijos. Sin embargo, en muchos casos la diferencia no alcanza la significación estadística, y en otros la magnitud de dicha diferencia no se considera relevante desde un punto de vista práctico. En cualquier caso, las mayores diferencias se observan cuando la segunda entrada la realiza un operador diferente y el patrón tiene un valor.

**Distribución de las discrepancias en función del entrevistador original:**

Comparación de la Fase A y la Fase B1: (datos en tantos por mil)

Al comparar las discrepancias en función del entrevistador original, el intervalo de confianza de la diferencia de proporciones se ha calculado respecto al entrevistador identificado como "1", ya que poseía mayor experiencia en el manejo y codificación manual de la entrevista.

Entrevistador	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
1	2.8261	–	40.2765	–
2	5.4491	1.57 a 3.68	61.1193	15.2 a 26.5
3	2.6567	N.S.	38.3387	N.S.
4	4.8962	1.06 a 3.08	50.9145	5.42 a 15.9

Entrevistador	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
1	11.3625	–	27.7396	–
2	2.9026	–10.8 a –6.14	35.0040	N.S.
3	1.9341	–11.6 a –7.22	26.1465	N.S.

Entrevistador	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
1	1.4516	-	22.2854	-
2	3.8299	1.53 a 3.22	38.9641	11.7 a 21.7
3	2.6482	0.29 a 2.10	24.6076	N.S.
4	1.6551	N.S.	35.9877	8.35 a 19.1
5	2.8073	N.S.	15.2937	-12.3 a -1.67
6	1.4993	N.S.	18.9873	N.S.
7	3.4671	N.S.	25.8064	N.S.

Entrevistador	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
1	2.2127	-	35.8083	-
2	1.8208	N.S.	39.8625	N.S.
3	1.4632	N.S.	37.5904	N.S.
4	1.7346	N.S.	33.6134	N.S.
5	1.5094	N.S.	27.8330	N.S.
6	1.7794	N.S.	48.2955	N.S.
7	3.1898	N.S.	56.9106	3.51 a 38.7

Comparación de la Fase A y la Fase B2: (en tantos por mil)

Entrevistador	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
1	5.9529	-	67.6099	-
2	8.9440	1.59 a 4.39	79.9889	5.63 a 19.1
3	5.7285	N.S.	66.3409	N.S.
4	7.3995	0.13 a 2.8	68.4627	N.S.

Los resultados obtenidos en relación a la distribución de las discrepancias según el entrevistador original indican que cuando el patrón está vacío las diferencias respecto al entrevistador 1 son, en su mayoría, no significativas o de magnitud no relevante desde un punto de vista práctico. En cambio, cuando el patrón está completo y aparecen discrepancias, éstas son más evidentes. En todos estos casos se observa que para el resto de entrevistadores se obtiene mayor proporción de discrepancias que para entrevistador 1, excepto en la condición No VA + ES. En esta última situación, para el entrevistador 5 se obtienen menos discrepancias.

**Distribución de las discrepancias en función del tipo de campo:**

Comparación de la Fase A y la Fase B1: (en tantos por mil)

Al comparar las discrepancias en función del tipo de campo, el intervalo de confianza de la diferencia de proporciones se ha calculado respecto a la categoría "campo numérico", ya que se ha considerado el más representativo cuando se registran datos psicodiagnósticos.

Campo	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Numérico	4.0819	–	46.4197	–
Tiempo	35.7143	N.S.	28.7770	N.S.
Fecha	0.0000	–4.48 a –3.68	66.0066	N.S.
Multirrespuesta	1.7637	N.S.	520.2020	404 a 543

Campo	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Numérico	5.3856	–	17.2816	–
Tiempo	250.0000	N.S.	0.0000	–19.7 a –14.9
Fecha	3.1601	N.S.	11.1111	N.S.
Multirrespuesta	5.2181	N.S.	327.8008	269 a 352

Campo	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Numérico	2.1802	–	19.7653	–
Tiempo	31.25	N.S.	43.3071	N.S.
Fecha	0.0000	N.S.	45.4545	1.5 a 49.9
Multirrespuesta	4.3196	0.77 a 3.51	245.8396	203 a 249

Campo	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Numérico	1.9032	–	38.1266	–
Tiempo	0.0000	–2.43 a –1.1	13.6986	N.S.
Fecha	0.0000	–	64.1026	N.S.
Multirrespuesta	0.0000	–2.43 a –1.38	139.5349	N.S.

### Comparación de la Fase A y la Fase B2:

Campo	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Numérico	7.1282	–	69.0293	–
Tiempo	107.1429	N.S.	57.5540	N.S.
Fecha	0.0000	–7.65 a 6.60	62.7063	N.S.
Multirrespuesta	1.7637	–8.86 a –1.87	500.0000	361 a 501

Los resultados obtenidos en relación a la distribución de las discrepancias según el tipo de campo indican que en los patrones sin valor las diferencias suelen ser no significativas o de magnitud despreciable desde un punto de vista práctico. En cambio, para los campos con valor en el patrón, los campos multirrespuesta acumulan mayor proporción de discrepancias respecto a los campos numéricos. La única excepción se evidencia en la condición No VA + ES, en cuyo caso la diferencia de proporciones no alcanza la significación estadística.

## Distribución de las discrepancias en función de la sección de la entrevista:

### Comparación de la Fase A y la Fase B1: (en tantos por mil)

Al comparar las discrepancias en función de la sección de la entrevista, el intervalo de confianza de la diferencia de proporciones se ha calculado respecto al apartado de "observaciones", ya que es el único que los entrevistadores codifican sin necesidad de formular ninguna pregunta a los informantes.

Sección	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Demográficos	16.7481	3.45 a 18.8	100.1106	52.2 a 79.2
TDAH	42.9777	25.7 a 49.0	16.7870	-27.6 a -7.72
Negativismo	8.2840	N.S.	31.9539	N.S.
T. Conducta	3.9560	N.S.	46.5611	N.S.
Alcohol	1.2256	-8.31 a -0.45	49.3730	N.S.
Tabaco	8.8832	N.S.	48.1651	N.S.
Colas	0.9524	-8.96 a -0.35	86.2069	9.05 a 94.5
Marihuana	0.7289	-8.80 a -0.96	30.6748	N.S.
Otras drogas	0.4148	-9.09 a -1.30	112.0879	47.2 a 108
Depresión	3.5872	N.S.	64.3330	18.5 a 41.3
Manía	1.2621	-8.48 a -0.22	26.3158	N.S.
Distimia	6.2053	N.S.	16.6920	-31.2 a -4.26
Ansiedad sep.	0.5168	-9.1 a -1.09	61.7284	9.54 a 45.0
T. evitación	1.2626	-8.6 a -0.10	35.7143	N.S.
Ansiedad exc.	7.9559	N.S.	26.7857	N.S.
Fobias	6.3851	N.S.	52.6316	5.62 a 30.8
TOC	18.6851	6.80 a 19.4	58.8235	9.43 a 39.3
Estrés postrau.	0.0000	-9.5 a -1.7	42.5532	N.S.
Anorexia	2.4651	N.S.	39.7946	N.S.
Bulimia	0.9709	-8.74 a -0.54	72.1649	10.4 a 65.1
Enuresis	1.4378	N.S.	66.8896	10.4 a 54.5
Encopresis	4.2662	N.S.	29.2683	N.S.
Menstruación	3.0628	N.S.	24.5283	N.S.
Id. sexual	3.5443	N.S.	71.8563	N.S.
Somatización	4.1786	N.S.	34.9599	N.S.
Psicosis	3.9336	N.S.	42.3223	N.S.
Estresores	14.6751	1.73 a 16.4	45.4780	N.S.
Observaciones	5.6101	-	34.4362	-



Sección	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Demográficos	4.2098	N.S.	119.1744	65.3 a 112
TDAH	5.5710	N.S.	11.1039	-36.8 a -2.33
Negativismo	9.5238	N.S.	8.3333	-40.3 a -4.38
T. Conducta	3.6731	-24.7 a -0.46	15.8730	N.S.
Alcohol	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Tabaco	47.8088	2.6 a 60.5	18.3486	N.S.
Colas	3.1546	N.S.	0.0000	-47.1 a -14.3
Marihuana	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Otras drogas	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Depresión	7.4359	N.S.	16.3522	N.S.
Manía	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Distimia	3.4247	N.S.	26.5487	N.S.
Ansiedad sep.	0.0000	-28.2 a -4.31	20.0803	N.S.
T. evitación	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Ansiedad exc.	23.0125	N.S.	3.0120	-45.1 a -10.2
Fobias	12.7226	N.S.	29.9625	N.S.
TOC	1.1765	-27.2 a -2.91	35.1648	N.S.
Estrés posttrau.	0.0000	-28.2 a -4.31	13.6054	N.S.
Anorexia	1.4556	-27.1 a -2.52	50.3876	N.S.
Bulimia	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Enuresis	0.0000	-28.2 a -4.31	87.5912	6.83 a 107
Encopresis	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Menstruación	0.0000	-28.2 a -4.31	20.1005	N.S.
Id. sexual	0.0000	-28.2 a -4.31	0.0000	-47.1 a -14.3
Somatización	16.1427	N.S.	14.1643	N.S.
Psicosis	52.4823	15.9 a 56.6	1.9608	-45.6 a -11.8
Estresores	7.7319	N.S.	11.6279	-37.5 a -0.52
Observaciones	16.2413	-	30.6604	-

Sección	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Demográficos	5.0500	-17.8 a -4.22	94.7181	71.7 a 92.5
TDAH	4.9614	-18.5 a -3.70	15.5057	N.S.
Negativismo	9.5694	N.S.	7.4799	N.S.
T. Conducta	1.8916	-20.9 a -7.47	18.6567	N.S.
Alcohol	0.0000	-22.7 a -9.41	19.7461	N.S.
Tabaco	0.0000	-22.7 a -9.41	7.7221	N.S.
Colas	0.0000	-22.7 a -9.41	22.0588	N.S.
Marihuana	0.2562	-22.5 a -9.14	5.6497	N.S.
Otras drogas	0.1104	-22.6 a -9.30	10.4712	N.S.
Depresión	3.3677	-19.4 a -5.97	28.9398	8.18 a 24.5
Manía	1.7475	-21.2 a -7.45	0.0000	-18.6 a -6.65
Distimia	0.0000	-22.7 a -9.41	13.3080	N.S.
Ansiedad sep.	8.6021	N.S.	14.0023	N.S.
T. evitación	0.6653	-22.2 a -8.62	56.3380	12.2 a 75.3
Ansiedad exc.	12.1507	N.S.	24.7845	0.53 a 23.8
Fobias	4.3158	-18.7 a -4.79	22.5244	1.46 a 18.4
TOC	2.1498	-20.8 a -7.4	6.6372	N.S.
Estrés postrau.	0.2823	-22.5 a -9.10	47.0219	10.4 a 58.4
Anorexia	0.4032	-22.4 a -8.96	49.7132	17.5 a 56.7
Bulimia	0.4843	-22.3 a -8.86	17.9372	N.S.
Enuresis	0.7252	-22.2 a -8.53	27.0833	N.S.
Encopresis	0.0000	-22.7 a -9.41	15.5440	N.S.
Menstruación	0.7358	-22.1 a -8.52	28.0112	N.S.
Id. sexual	1.6411	-21.3 a -7.51	5.7471	N.S.
Somatización	1.0959	-21.7 a -8.23	42.9043	17.4 a 43.2
Psicosis	2.7726	-20.3 a -6.28	5.3035	-14.2 a -0.421
Estresores	6.6865	-17.3 a -1.43	18.8889	N.S.
Observaciones	16.0701	-	12.6113	-

Sección	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Demográficos	5.2083	–	57.5858	21.3 a 60.4
TDAH	1.8382	N.S.	17.3675	N.S.
Negativismo	16.5975	N.S.	26.0304	N.S.
T. Conducta	2.3522	N.S.	44.9102	10.4 a 45.9
Alcohol	0.0000	N.S.	17.8571	N.S.
Tabaco	0.0000	N.S.	62.5000	N.S.
Colas	0.0000	N.S.	324.3243	156 a 459
Marihuana	0.4581	N.S.	0.0000	–29 a –4.4
Otras drogas	0.0000	N.S.	210.5263	101 a 286
Depresión	2.3280	N.S.	35.8744	1.80 a 36.5
Manía	1.6313	N.S.	4.8544	N.S.
Distimia	4.9669	N.S.	20.4082	N.S.
Ansiedad sep.	0.0000	N.S.	38.0435	N.S.
T. evitación	0.0000	N.S.	0.0000	–29 a –4.4
Ansiedad exc.	1.9685	N.S.	0.0000	–29 a –4.4
Fobias	6.0180	N.S.	44.5063	8.3 a 47.2
TOC	7.9365	N.S.	18.6667	N.S.
Estrés postrau.	1.0091	N.S.	0.0000	–29 a –4.4
Anorexia	2.7933	N.S.	38.8349	N.S.
Bulimia	0.0000	N.S.	58.8236	N.S.
Enuresis	2.8329	N.S.	25.9740	N.S.
Encopresis	0.0000	N.S.	42.8571	N.S.
Menstruación	0.0000	N.S.	0.0000	–29 a –4.4
Id. sexual	1.9724	N.S.	128.2051	5.81 a 217
Somatización	0.0000	N.S.	21.3523	N.S.
Psicosis	0.0000	N.S.	91.9037	46 a 104
Estresores	8.7209	N.S.	36.9650	N.S.
Observaciones	6.1920	–	16.7464	–

Comparación de la Fase A y la Fase B2: (en tantos por mil)

Sección	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Discrepancias	IC 95% $\delta$	Discrepancias	IC 95% $\delta$
Demográficos	18.8416	N.S.	146.0177	17 a 56.3
TDAH	43.7452	15.5 a 41.1	24.0094	-102 a -68.9
Negativismo	14.2012	N.S.	46.6213	-81.3 a -44.3
T. Conducta	12.1319	N.S.	95.9584	N.S.
Alcohol	2.0135	-19.9 a -6.96	43.8871	-85 a -46
Tabaco	8.8832	N.S.	61.9266	-75.1 a -19.8
Colas	1.9048	-20.4 a -6.6	609.1954	426 a 574
Marihuana	2.6725	-19.2 a -6.3	49.0798	-85.2 a -35.4
Otras drogas	0.2074	-21.6 a -8.82	43.9560	-90.1 a -40.8
Depresión	7.8469	-14.2 a -1.01	102.1760	N.S.
Manía	2.1035	-20.0 a -6.67	25.1196	-103 a -65.2
Distimia	1.9093	-20.2 a -6.85	21.2443	-107 a -68.8
Ansiedad sep.	1.5504	-20.5 a -7.24	50.4115	-80 a -38
T. evitación	0.6313	-21.3 a -8.28	39.6825	-98.6 a -40.8
Ansiedad exc.	6.1199	-16.7 a -1.88	50.8929	-78.9 a -38
Fobias	15.2259	N.S.	51.5038	-75.9 a -39.9
TOC	20.4152	N.S.	63.3484	-66 a -26
Estrés postrau.	8.6299	N.S.	35.4610	-97.7 a -50.2
Anorexia	8.2169	N.S.	61.6175	-71.0 a -24.6
Bulimia	0.9709	-21.0 a -7.92	79.8969	N.S.
Enuresis	0.0000	-21.8 a -9.03	90.3010	N.S.
Encopresis	4.2662	-18.6 a -3.8	58.5366	-86.7 a -15.0
Menstruación	3.0628	-19.4 a -5.3	32.0755	-99.2 a -55.5
Id. sexual	6.0759	-16.6 a -2.1	59.8802	-88.8 a -10.2
Somatización	7.8349	-14.6 a -0.6	43.7000	-84.9 a -46.5
Psicosis	7.4301	-15.3 a -0.7	59.6853	-68.9 a -30.5
Estresores	16.0727	N.S.	36.1757	-91.2 a -55.3
Observaciones	15.4278	-	109.3856	-

Los resultados obtenidos en relación a la distribución de discrepancias según la sección de la entrevista indican que las mayores diferencias respecto al apartado de observaciones se obtienen cuando el patrón tiene un valor. En este caso, sin embargo, es difícil determinar si existen secciones que sistemáticamente acumulen mayor proporción de discrepancias, ya que los resultados fluctúan notablemente según la condición de registro. Estos resultados sugieren que no existe una sección concreta que favorezca que los operadores de entrada de datos cometan mayor proporción de discrepancias. La única excepción es demográficos, para la cual siempre se obtienen proporciones de discrepancias manifiestamente mayores, independientemente de la condición de registro.

### 4.3.2. Estudio de los errores entre los datos verificados y el patrón

#### Errores totales en función de la condición de registro:

La proporción total de errores (en tantos por mil) para cada condición de registro vale (al comparar la Fase A con el patrón):

Condición registro	Si patrón está vacío	Si patrón tiene un valor	TOTAL
ES + DE (misma operadora)	4.1115	38.4285	13.6770
ES + ES	4.3120	30.6598	11.5972
NO VA + ES	2.3379	29.0477	8.8924

Estos resultados indican que la condición de ES para la Fase A (n=153 y n=45) obtiene mayor proporción de errores que la fase en que se activa la verificación aleatoria (No VA + ES).

La proporción total de errores (en tantos por mil) para cada condición de registro vale (al comparar la Fase B verificada con el patrón):

Condición registro	Si patrón está vacío	Si patrón tiene un valor	TOTAL
DE (misma operadora)	2.1664	24.7841	8.4708
ES + ES	0.5624	8.0124	2.6224
NO VA + ES	0.5751	7.7692	2.3405
VA	2.0095	24.7326	7.3297
DE (operadores distintos)	1.1615	6.4405	2.6330

Estos resultados indican que la mayor proporción de errores corresponde a la condición DE realizada por la misma operadora, seguida de los registros que habían sido verificados por el propio Sistema DAT al activar la verificación aleatoria.

La tabla anterior también indica que cuando se compara la Fase B verificada con el patrón la proporción de error es notablemente menor que cuando se compara la Fase A con el patrón, independientemente de la condición de registro.

#### Distribución de los errores en función del tipo de protocolo:

Comparación de la Fase A y el patrón: (en tantos por mil)

Tipo protocolo	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	2.7980	–	26.4322	–
Padres	4.9016	1.37 a 2.84	48.5839	18.6 a 25.8

Tipo protocolo	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	1.9300	–	26.3554	–
Padres	5.6297	2.40 a 5.00	33.9644	1.54 a 13.7

Tipo protocolo	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	1.5740	–	15.4625	–
Padres	2.7555	0.62 a 1.74	39.0362	20.2 a 26.9

Los resultados obtenidos en relación a la distribución de los errores según el tipo de protocolo indican que al comparar la Fase A con el patrón las entrevistas de los padres acumulan mayor proporción de errores que las entrevistas de los hijos, independientemente de la condición de registro. En cualquier caso, cuando el patrón está vacío las diferencias son poco relevantes desde un punto de vista práctico, y cuando el patrón tiene un valor dichas diferencias se hacen más importantes.

Comparación de la Fase B1 verificada y el patrón: (en tantos por mil)

Tipo protocolo	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	1.2763	–	14.1006	–
Padres	2.7018	0.90 a 1.95	33.8281	16.8 a 22.6

Tipo protocolo	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	0.4386	–	4.3298	–
Padres	0.6309	N.S.	10.8397	3.5 a 9.5

Tipo protocolo	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	0.2668	–	3.5577	–
Padres	0.7435	0.21 a 0.74	10.8657	5.57 a 9.04

Tipo protocolo	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	1.4051	–	13.5766	–
Padres	2.6361	0.20 a 2.26	37.0718	16.9 a 30.1

Comparación de la Fase B2 verificada y el patrón: (en tantos por mil)

Tipo protocolo	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Hijos	0.4418	–	3.1739	–
Padres	1.5945	0.79 a 1.52	9.2058	4.56 a 7.51

Al igual que en el caso anterior, al comparar la Fase B verificada con el patrón las entrevistas de los padres acumulan mayor proporción de errores, independientemente de la condición de registro. En cualquier caso, es importante notar que cuando el patrón está vacío dichas diferencias son poco relevantes desde el punto de vista práctico, y también que independientemente del contenido del patrón las diferencias entre las proporciones son menores que al comparar la Fase A con el patrón.

### Distribución de los errores en función del entrevistador original:

#### Comparación de la Fase A y el patrón: (en tantos por mil)

Entrevistador	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	3.1917	–	34.2027	–
2	4.9835	0.80 a 2.80	45.1504	5.98 a 15.9
3	3.3298	N.S.	30.2032	N.S.
4	4.7279	0.56 a 2.51	40.3342	1.46 a 10.8

Entrevistador	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	9.6828	–	31.5847	–
2	2.4022	–9.42 a –5.14	33.9433	N.S.
3	1.3454	–10.4 a –6.32	27.3916	N.S.

Entrevistador	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	1.4214	–	19.8311	–
2	3.9787	1.70 a 3.41	37.9046	13.2 a 23
3	1.8086	N.S.	24.2201	N.S.
4	1.8728	N.S.	38.2848	13.1 a 23.8
5	3.4089	0.32 a 3.66	24.6785	N.S.
6	1.4993	N.S.	18.9873	N.S.
7	2.4765	N.S.	27.9570	N.S.

Los resultados obtenidos en relación a la distribución de los errores según el entrevistador original indican que al comparar la Fase A con el patrón la mayor proporción de errores se acumula cuando el patrón tiene un valor, y en concreto para los entrevistadores 2 y 4 (en relación al entrevistador 1). Sin embargo, en la condición de registro ES + ES, esta diferencia no alcanza la significación estadística.

#### Comparación de la Fase B1 verificada y el patrón: (en tantos por mil)

Entrevistador	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	1.8985	–	13.7436	–
2	2.0278	N.S.	40.6889	22.8 a 31.1
3	2.5730	N.S.	21.7829	3.67 a 12.4
4	2.4478	N.S.	23.1400	6.07 a 12.7

Entrevistador	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	0.9880	-	10.9085	-
2	0.6005	N.S.	6.0992	-11.2 a -2.38
3	0.1682	-1.47 a -0.2	5.8103	-11.3 a -2.85

Entrevistador	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	0.3932	-	7.0685	-
2	0.8552	N.S.	6.3567	N.S.
3	0.1938	N.S.	6.5879	N.S.
4	0.6098	N.S.	9.9541	N.S.
5	1.4036	N.S.	13.2082	1.66 a 10.6
6	0.0000	N.S.	3.1646	N.S.
7	0.4953	N.S.	4.3011	N.S.

Entrevistador	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	2.0352	-	45.7317	-
2	2.0442	N.S.	20.0000	-37.4 a -14.1
3	0.8944	N.S.	13.0902	-41.3 a -22.2
4	1.7575	N.S.	15.5763	-41.2 a -19.1
5	4.8309	N.S.	25.1451	-36.9 a -4.24
6	1.6026	N.S.	11.1421	-48.8 a -20.3
7	4.3668	N.S.	36.0360	N.S.

Comparación de la Fase B2 verificada y el patrón: (en tantos por mil)

Entrevistador	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
1	0.9080	-	6.0909	-
2	2.0621	0.55 a 1.76	7.4953	N.S.
3	0.5497	N.S.	5.8576	N.S.
4	0.8718	N.S.	6.1064	N.S.

Los resultados obtenidos en relación a la distribución de los errores según el entrevistador original indican que al comparar la Fase B verificada con el patrón las diferencias entre el entrevistador 1 y el resto de entrevistadores no son estadísticamente significativas o prácticamente importantes cuando el patrón está vacío, y tampoco cuando patrón tiene un valor y la condición de registro es ES + DE con diferente operador. En el resto de situaciones, sin embargo, las discrepancias son más importantes, aunque no siguen una línea sistemáticamente consistente en las diferentes condiciones de registro.



## Distribución de los errores en función del tipo de campo:

### Comparación de la Fase A y el patrón: (en tantos por mil)

Campo	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	3.7570	–	27.3116	–
Tiempo	0.0000	–4.14 a –3.37	25.1798	N.S.
Fecha	0.0000	–4.14 a –3.37	39.6040	N.S.
Multirrespuesta	7.6275	2.12 a 5.62	368.1885	315 a 367

Campo	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	4.0477	–	16.1583	–
Tiempo	0.0000	–4.78 a –3.32	0.0000	–18.5 a –13.9
Fecha	0.0000	–4.78 a –3.32	0.0000	–18.5 a –13.9
Multirrespuesta	7.0225	N.S.	390.0415	330 a 417

Campo	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	2.0252	–	19.1217	–
Tiempo	0.0000	–2.31 a –1.74	27.5591	N.S.
Fecha	0.0000	–2.31 a –1.74	31.4685	N.S.
Multirrespuesta	5.6156	2.04 a 5.14	273.8275	231 a 279

Los resultados obtenidos en relación a la distribución de los errores según el tipo de campo indican que al comparar la Fase A con el patrón la mayor proporción de errores se acumula cuando el patrón tiene un valor, y en particular al comparar los campos multirrespuesta con los numéricos.

### Comparación de la Fase B1 verificada y el patrón: (en tantos por mil)

Campo	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	1.4114	–	17.1072	–
Tiempo	0.0000	–1.65 a –1.18	21.5827	N.S.
Fecha	0.0000	–1.65 a –1.18	16.5016	N.S.
Multirrespuesta	9.6347	6.29 a 10.2	253.3137	213 a 259

Campo	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	0.1029	–	4.6660	–
Tiempo	0.0000	N.S.	0.0000	–5.91 a –3.42
Fecha	0.0000	N.S.	0.0000	–5.91 a –3.42
Multirrespuesta	5.2669	2.50 a 7.82	91.2863	60.9 a 112

Campo	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	0.4133	–	6.2513	–
Tiempo	0.0000	–0.54 a –0.28	0.0000	–7.11 a –5.40
Fecha	0.0000	–0.54 a –0.28	3.4965	N.S.
Multirrespuesta	2.2678	0.88 a 2.83	47.6551	29.9 a 52.9

Campo	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	1.6793	–	19.2987	–
Tiempo	0.0000	–2.17 a –1.19	27.3973	N.S.
Fecha	0.0000	–2.17 a –1.19	25.6410	N.S.
Multirrespuesta	5.4709	0.89 a 6.69	165.1376	105 a 186

Comparación de la Fase B2 verificada y el patrón: (en tantos por mil)

Campo	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Numérico	1.1880	–	4.1768	–
Tiempo	71.4286	N.S.	3.5971	N.S.
Fecha	0.0000	–1.4 a –0.97	9.9010	N.S.
Multirrespuesta	0.7025	N.S.	72.9013	54.9 a 82.6

Al igual que en el caso anterior, los resultados obtenidos en relación a la distribución de los errores según el tipo de campo indican que al comparar la Fase B verificada con el patrón la mayor proporción de errores se acumula cuando el patrón tiene un valor, y en concreto cuando los campos son del tipo multirrespuesta.

**Distribución de los errores en función de la sección de la entrevista:**

Comparación de la Fase A y el patrón: (en tantos por mil)

Sección	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	13.4811	N.S.	153.8945	106 a 133
TDAH	45.2801	19.2 a 44.7	15.8110	-28.5 a -8.73
Negativismo	4.7337	-16.1 a -1.05	11.0005	-33.8 a -13.0
T. Conducta	3.6923	-15.7 a -3.58	21.9806	-22.7 a -2.17
Alcohol	0.5253	-18.8 a -6.83	21.9436	-24.8 a -0.21
Tabaco	8.8832	N.S.	18.3486	-31.7 a -0.44
Colas	0.9524	-18.6 a -6.13	28.7356	N.S.
Marihuana	0.2430	-19.0 a -7.12	10.2249	-37.1 a -11.3
Otras drogas	0.1037	-19.2 a -7.27	15.3846	-33.7 a -4.42
Depresión	2.7651	-16.6 a -4.54	44.0870	N.S.
Manía	0.4207	-18.9 a -6.9	8.3732	-37.2 a -14.9
Distimia	0.4773	-18.9 a -6.82	7.5873	-38.3 a -15.4
Ansiedad sep.	0.0000	-19.3 a -7.37	33.9506	N.S.
T. evitación	0.6313	-18.8 a -6.61	11.9048	-38.8 a -6.24
Ansiedad exc.	5.5080	-14.8 a -0.87	16.0714	-30.2 a -6.51
Fobias	4.4204	-15.2 a -2.61	24.8120	N.S.
TOC	15.5709	N.S.	50.4202	1.66 a 30.3
Estrés posttrau.	0.0000	-19.3 a -7.37	16.5485	-33.2 a -2.59
Anorexia	1.6434	-17.8 a -5.52	11.5533	-34.8 a -10.9
Bulimia	0.0000	-19.3 a -7.37	64.4330	3.86 a 56.1
Enuresis	0.7189	-18.7 a -6.49	16.7224	-31.6 a -3.87
Encopresis	4.2662	-16.1 a -2.03	9.7561	-41.0 a -8.33
Menstruación	0.7657	-18.7 a -6.42	7.5472	-38.7 a -15.0
Id. sexual	4.0506	-15.9 a -2.70	47.9042	N.S.
Somatización	3.6563	-16 a -3.52	15.2950	-30.5 a -7.81
Psicosis	1.3112	-18.1 a -5.88	5.9685	-38.4 a -18.5
Estresores	7.6869	N.S.	19.1214	-26.4 a -4.20
Observaciones	13.3240	-	34.4362	-

Sección	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	8.0959	N.S.	159.121	115 a 161
TDAH	5.5710	N.S.	9.7975	N.S.
Negativismo	4.7619	N.S.	6.6667	N.S.
T. Conducta	2.7548	N.S.	9.8262	N.S.
Alcohol	0.2793	N.S.	0.0000	-34.9 a -7.51
Tabaco	0.0000	N.S.	0.0000	-34.9 a -7.51
Colas	3.1546	N.S.	0.0000	-34.9 a -7.51
Marihuana	0.0000	N.S.	0.0000	-34.9 a -7.51
Otras drogas	0.0000	N.S.	0.0000	-34.9 a -7.51
Depresión	7.4359	N.S.	17.6101	N.S.
Manía	0.0000	N.S.	0.0000	-34.9 a -7.51
Distimia	3.4247	N.S.	0.0000	-34.9 a -7.51
Ansiedad sep.	0.0000	N.S.	8.0321	N.S.
T. evitación	0.0000	N.S.	0.0000	-34.9 a -7.51
Ansiedad exc.	0.0000	N.S.	0.0000	-34.9 a -7.51
Fobias	4.2409	N.S.	21.2235	N.S.
TOC	1.1765	N.S.	35.1648	N.S.
Estrés postrau.	0.0000	N.S.	13.6054	N.S.
Anorexia	1.4556	N.S.	27.1318	N.S.
Bulimia	0.0000	N.S.	90.4167	N.S.
Enuresis	0.0000	N.S.	65.6934	0.77 a 88.2
Encopresis	0.0000	N.S.	33.8983	N.S.
Menstruación	0.0000	N.S.	15.0754	N.S.
Id. sexual	0.0000	N.S.	0.0000	-34.9 a -7.51
Somatización	16.1427	1.86 a 21.1	5.6657	N.S.
Psicosis	52.4823	30.2 a 65.5	1.9608	-33.5 a -5.02
Estresores	7.7320	N.S.	8.3057	N.S.
Observaciones	4.6404	-	21.2264	-

Sección	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	8.1196	-15.7 a -1.65	130.5995	106 a 130
TDAH	4.9614	-19.4 a -4.30	13.1202	N.S.
Negativismo	3.5885	-21.1 a -5.29	8.0552	N.S.
T. Conducta	1.8916	-21.8 a -8.05	14.6383	N.S.
Alcohol	0.0896	-22.6 a -9.32	8.4626	N.S.
Tabaco	1.1299	-22.8 a -8.51	11.5830	N.S.
Colas	0.0000	-23.6 a -9.99	7.3529	N.S.
Marihuana	0.2562	-23.4 a -9.73	20.2899	N.S.
Otras drogas	0.0000	-23.6 a -9.99	10.4712	N.S.
Depresión	2.6512	-21.0 a -7.29	19.7708	N.S.
Manía	0.0000	-23.6 a -9.99	0.0000	-18.6 a -6.65
Distimia	0.0000	-23.6 a -9.99	13.3080	N.S.
Ansiedad sep.	9.1398	-15.7 a -0.40	10.5017	N.S.
T. evitación	0.0000	-23.6 a -9.99	56.3380	12.2 a 75.3
Ansiedad exc.	12.1507	-13.3 a -3.97	10.7759	N.S.
Fobias	2.7926	-21.0 a -7.0	16.9996	N.S.
TOC	1.4332	-22.3 a -8.42	5.1622	-14.5 a -0.37
Estrés postrau.	0.2823	-23.3 a -9.69	31.3480	N.S.
Anorexia	0.4032	-23.3 a -9.54	47.8011	16.0 a 54.4
Bulimia	0.4843	-23.2 a -9.44	13.4529	N.S.
Enuresis	0.7252	-23.0 a -9.12	16.6667	N.S.
Encopresis	0.0000	-23.6 a -9.99	5.1813	N.S.
Menstruación	0.0000	-23.6 a -9.99	22.4090	N.S.
Id. sexual	0.5470	-21.3 a -9.36	0.0000	-18.6 a -6.65
Somatización	0.8219	-22.4 a -8.91	23.1023	0.146 a 20.8
Psicosis	0.4621	-23.2 a -9.47	1.7678	-17.1 a -4.56
Estresores	5.9435	-18.8 a -2.91	10.0000	N.S.
Observaciones	16.8006	-	12.6113	-

Los resultados obtenidos al comparar la Fase A con el patrón en función de la sección de la entrevista indican que las mayores diferencias se acumulan cuando el patrón tiene un valor, y en concreto para la sección de demográficos.

Comparación de la Fase B1 verificada y el patrón: (en tantos por mil)

Sección	Identificador del registro: ES + DE (misma operadora)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	14.4968	N.S.	105.3183	67.1 a 90.9
TDAH	17.6516	N.S.	8.1983	-26.7 a -9.61
Negativismo	1.1834	-14.3 a -3.02	8.3814	-27.1 a -8.83
T. Conducta	1.4066	-13.6 a -3.25	11.5812	-23.5 a -5.98
Alcohol	0.0000	-14.9 a -4.70	3.1348	-31.9 a -14.5
Tabaco	0.0000	-14.9 a -4.70	2.2936	-33.3 a -14.7
Colas	0.0000	-14.9 a -4.70	17.2414	-30.1 a -11.9
Marihuana	0.0000	-14.9 a -4.70	8.1799	-29.6 a -6.74
Otras drogas	0.1037	-14.8 a -4.59	6.5934	-30.8 a -8.7
Depresión	0.3737	-14.6 a -4.32	30.8420	N.S.
Manía	0.4207	-14.6 a -4.21	5.9809	-30.0 a -10.7
Distimia	0.0000	-14.9 a -4.70	6.0698	-30.3 a -10.2
Ansiedad sep.	0.0000	-14.9 a -4.70	31.8930	N.S.
T. evitación	0.0000	-14.9 a -4.70	3.9683	-33.6 a -11.1
Ansiedad exc.	0.0000	-14.9 a -4.70	8.9286	-27.2 a -7.56
Fobias	0.4912	-14.5 a -4.16	15.4135	-20.3 a -1.52
TOC	1.0381	-14.0 a -3.53	21.3316	N.S.
Estrés postrau.	0.0000	-14.9 a -4.70	2.3641	-33.3 a -14.6
Anorexia	0.0000	-14.9 a -4.70	2.5674	-32.7 a -14.9
Bulimia	0.0000	-14.9 a -4.70	64.4330	12.3 a 63.9
Enuresis	0.7189	-14.4 a -3.79	15.0502	N.S.
Encopresis	0.0000	-14.9 a -4.70	9.7561	-32.3 a -0.84
Menstruación	0.7657	-14.4 a -3.72	0.0000	-34.5 a -18.2
Id. sexual	0.4937	-13.3 a -2.30	23.9521	N.S.
Somatización	0.7835	-14.2 a -3.84	5.8267	-29.6 a -11.4
Psicosis	0.0000	-14.9 a -4.70	7.5963	-27.8 a -9.67
Estresores	2.0964	-13.4 a -2.08	11.3695	-24.4 a -5.54
Observaciones	9.8177	-	26.3336	-

Sección	Identificador del registro: ES + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	5.5052	2.90 a 8.11	54.5939	43.1 a 66.1
TDAH	0.0000	–	1.9595	N.S.
Negativismo	0.0000	–	0.0000	–
T. Conducta	0.0000	–	0.7559	N.S.
Alcohol	0.2793	N.S.	0.0000	–
Tabaco	0.0000	–	0.0000	–
Colas	0.0000	–	0.0000	–
Marihuana	0.0000	–	0.0000	–
Otras drogas	0.0000	–	0.0000	–
Depresión	0.0000	–	3.7736	N.S.
Manía	0.0000	–	0.0000	–
Distimia	0.0000	–	0.0000	–
Ansiedad sep.	0.0000	–	0.0000	–
T. evitación	0.0000	–	0.0000	–
Ansiedad exc.	0.0000	–	0.0000	–
Fobias	0.0000	–	1.2484	N.S.
TOC	0.0000	–	0.0000	–
Estrés postrau.	0.0000	–	0.0000	–
Anorexia	0.0000	–	0.0000	–
Bulimia	0.0000	–	10.4167	N.S.
Enuresis	0.0000	–	0.0000	–
Encopresis	0.0000	–	33.8983	N.S.
Menstruación	0.0000	–	5.0251	N.S.
Id. sexual	0.0000	–	0.0000	–
Somatización	0.0000	–	0.0000	–
Psicosis	0.0000	–	0.0000	–
Estresores	0.0000	–	1.6611	N.S.
Observaciones	0.0000	–	0.0000	–

Sección	Identificador del registro: No VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	4.2578	1.61 a 5.44	48.8077	38.9 a 52.8
TDAH	0.5513	N.S.	1.6698	N.S.
Negativismo	0.0000	N.S.	1.7261	N.S.
T. Conducta	0.1801	N.S.	2.5832	N.S.
Alcohol	0.0896	N.S.	1.4104	N.S.
Tabaco	1.1299	N.S.	7.7220	N.S.
Colas	0.0000	N.S.	0.0000	-5.87 a -0.064
Marihuana	0.0000	N.S.	17.4000	0.33 a 28.5
Otras drogas	0.0000	N.S.	2.6178	N.S.
Depresión	0.5016	N.S.	1.7192	N.S.
Manía	0.0000	N.S.	0.0000	-5.87 a -0.064
Distimia	0.0000	N.S.	0.0000	-5.87 a -0.064
Ansiedad sep.	0.5376	N.S.	0.0000	-5.87 a -0.064
T. evitación	0.0000	N.S.	0.0000	-5.87 a -0.064
Ansiedad exc.	0.6075	N.S.	0.0000	-5.87 a -0.064
Fobias	0.5070	N.S.	1.2750	N.S.
TOC	0.0000	N.S.	0.7375	N.S.
Estrés postrau.	0.0000	N.S.	0.0000	-5.87 a -0.064
Anorexia	0.0000	N.S.	1.9120	N.S.
Bulimia	0.0000	N.S.	0.0000	-5.87 a -0.064
Enuresis	0.0000	N.S.	2.0833	N.S.
Encopresis	0.0000	N.S.	0.0000	-5.87 a -0.064
Menstruación	0.0000	N.S.	0.0000	-5.87 a -0.064
Id. sexual	0.0000	N.S.	0.0000	-5.87 a -0.064
Somatización	0.2740	N.S.	0.8251	N.S.
Psicosis	0.0000	N.S.	0.0000	-5.87 a -0.064
Estresores	0.0000	N.S.	1.6667	N.S.
Observaciones	0.7305	-	2.9674	-



Sección	Identificador del registro: VA + ES			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	6.4493	3.48 a 9.42	77.5063	53.1 a 87.6
TDAH	14.7058	4.59 a 24.8	15.5393	N.S.
Negativismo	0.0290	N.S.	41.2148	14.2 a 53.9
T. Conducta	0.3360	N.S.	8.9820	N.S.
Alcohol	0.0000	–	17.8571	N.S.
Tabaco	0.0000	–	0.0000	N.S.
Colas	0.0000	–	0.0000	N.S.
Marihuana	0.0000	–	0.0000	N.S.
Otras drogas	0.0000	–	0.0000	N.S.
Depresión	0.5173	N.S.	43.7220	20.9 a 52.2
Manía	0.0000	–	0.0000	N.S.
Distimia	0.0000	–	0.0000	N.S.
Ansiedad sep.	0.0000	–	5.4348	N.S.
T. evitación	0.0000	–	0.0000	N.S.
Ansiedad exc.	0.0000	–	82.4742	35.8 a 11.5
Fobias	9.0271	3.16 a 14.9	2.7882	N.S.
TOC	1.3228	N.S.	2.6667	N.S.
Estrés postrau.	1.0091	N.S.	0.0000	N.S.
Anorexia	2.7933	N.S.	9.7087	N.S.
Bulimia	0.0000	–	0.0000	N.S.
Enuresis	19.8300	5.29 a 3.44	12.9870	N.S.
Encopresis	0.0000	–	57.1429	N.S.
Menstruación	0.0000	–	21.2766	N.S.
Id. sexual	1.9724	N.S.	0.0000	N.S.
Somatización	0.0000	–	14.2349	N.S.
Psicosis	0.0000	–	10.9409	N.S.
Estresores	0.1720	N.S.	7.7821	N.S.
Observaciones	0.0000	–	7.1770	–

Comparación de la Fase B2 verificada y el patrón: (en tantos por mil)

Sección	Identificador del registro: ES + DE (diferente operador)			
	Patrón vacío		Patrón con valor	
	Errores	IC 95% $\delta$	Errores	IC 95% $\delta$
Demográficos	6.3712	N.S.	38.5260	28.1 a 40.8
TDAH	17.6516	3.91 a 20.2	4.4896	N.S.
Negativismo	1.1834	N.S.	1.5715	N.S.
T. Conducta	0.8791	-8.65 a -0.80	1.4181	N.S.
Alcohol	0.0000	-9.49 a -1.73	5.4859	N.S.
Tabaco	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Colas	0.9524	-8.96 a -0.35	0.0000	-7.3 a -0.82
Marihuana	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Otras drogas	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Depresión	0.4484	-9.05 a -1.30	2.6490	N.S.
Manía	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Distimia	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Ansiedad sep.	0.0000	-9.49 a -1.73	3.0864	N.S.
T. evitación	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Ansiedad exc.	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Fobias	0.2456	-9.27 a -1.46	0.7519	N.S.
TOC	0.6920	-8.91 a -0.92	3.2321	N.S.
Estrés postrau.	0.0000	-9.49 a -1.73	0.0000	-7.3 a -0.82
Anorexia	0.4108	-9.16 a -1.24	1.1837	N.S.
Bulimia	0.0000	-9.49 a -1.73	7.7320	N.S.
Enuresis	0.0000	-9.49 a -1.73	1.6722	N.S.
Encopresis	0.0000	-9.49 a -1.73	9.7561	N.S.
Menstruación	0.0000	-9.49 a -1.73	1.8868	N.S.
Id. sexual	1.5190	N.S.	17.9641	N.S.
Somatización	0.2612	-9.26 a -1.44	0.7283	N.S.
Psicosis	0.0000	-9.49 a -1.73	1.0852	N.S.
Estresores	0.0000	-9.49 a -1.73	1.5504	N.S.
Observaciones	5.6101	-	4.0513	-

Los resultados obtenidos al comparar la Fase B verificada con el patrón en función de la sección de la entrevista indican que las mayores diferencias se acumulan cuando el patrón tiene un valor, y en concreto para la sección de demográficos. Estas proporciones de error, sin embargo, son inferiores que cuando se compara la Fase A con el patrón.

#### 4.4. DISCUSIÓN

Los resultados presentados en el apartado anterior sugieren que la verificación aleatoria de un porcentaje de los registros contribuye a incrementar la eficacia global del proceso de captura de datos, en concreto cuando dicha estrategia se compara con otras técnicas tales como la doble entrada del total de los registros. Esto se evidencia cuando:

1. Se valoran las discrepancias que comete un operador que entra dos veces la misma serie de datos (primer objetivo del estudio). En este caso se ha comprobado que la condición más favorable para el registro sin discordancias corresponde al momento en que se activa la verificación aleatoria. En estos casos la proporción total de discrepancias (en tantos por mil) es igual a 8.9 para la condición "No VA + ES" y 10.9 para la condición "VA". La mayor proporción de discrepancias se obtiene cuando la doble entrada la realiza un operador distinto para el 100% de los registros (25.7).
2. Se valoran los errores que persisten cuando se compara la Fase A con el patrón (segundo objetivo del estudio). En este caso también se evidencia que la proporción total de errores más pequeña corresponde a la condición "No VA + ES" (8.9). En cambio, la doble entrada sin "feedback" obtiene una proporción de errores igual a 11.6, y la doble entrada con "feedback" una proporción igual a 13.7.

Sin embargo, la verificación aleatoria no resulta más eficaz que la doble entrada cuando se valoran los errores que persisten al comparar la Fase B verificada con el patrón. En este caso, los resultados del estudio indican que las proporciones de error totales son similares en las condiciones de registro correspondientes a la doble entrada con diferentes operadores (2.6), la doble entrada sin "feedback" (2.6) y la verificación aleatoria (2.3). Estos resultados sugieren que las condiciones más óptimas de registro se dan en el caso de que se efectúe una doble entrada con diferentes interpretaciones de la información que se captura, y cuando un único operador dispone de información sobre el control que se está ejerciendo sobre la grabación de datos y se le solicita aleatoriamente que vuelva a grabar algunas entrevistas.

En conjunto, estos resultados sugieren que la verificación aleatoria constituye un procedimiento adecuado como técnica de control de la calidad de la captura de datos clínicos. Es importante destacar que la principal ventaja de la verificación aleatoria consiste en que reduce considerablemente el coste total del registro de datos (y por extensión de la investigación) frente a la doble entrada. También es importante tener presente que ambas estrategias constituyen sistemas de control de la calidad de datos compatibles. En este sentido, por ejemplo, se puede utilizar la verificación aleatoria durante la captura de los datos y completar la grabación con la doble entrada del resto de los registros. Estas técnicas también se pueden utilizar simultáneamente en estudios con operadores o registros que requieran diferentes niveles de control.

Los resultados de este estudio también indican que existen variables que contribuyen a decrementar la eficiencia del registro de datos. En concreto, la experiencia que se ha realizado sugiere que los factores que incrementan el proporción de discrepancias y errores son: (1) el diseño del protocolo donde se registran manualmente los datos (las hojas de registro para padres obtienen mayor proporción de discrepancias-errores que las hojas para hijos); (2) el tipo de campo (las multirrespuestas son las variables más desfavorables para el registro); y (3) la disponibilidad de entrevistas codificadas a mano por diferentes entrevistadores. En relación a estas conclusiones nos gustaría destacar que:

1. El diseño del protocolo del registro de datos es fundamental, ya que puede actuar dificultando el registro de la información para el entrevistador que codifica a mano la información y para el operador que captura dichos datos en formato informatizado. En particular, es probable que en este estudio las hojas de registro para padres hayan

incrementado la proporción de discordancias y errores debido a que en una misma hoja se registra información de varios registros (en concreto, de hasta tres hijos).

2. En cuanto a la dificultad evidenciada por los campos multirrespuesta, es importante destacar que en el formulario que hemos preparado para recoger datos de la entrevista DICA-III-R se han definido 3 campos de este tipo: cursos repetidos por el sujeto, cursos saltados y organigrama familiar. En particular, debido a la gran acumulación de discrepancias/errores en este tipo de campo, hemos comprobado que en su mayoría dichas discordancias corresponden al organigrama familiar. Consideramos que en este caso, las dificultades del registro pueden deberse tanto al tipo de campo como al requerimiento de que el/la operador/a interprete dichos contenidos (ver Anexos 1A y 1B). En efecto, en última instancia, los códigos que se graban dentro de los campos multirrespuesta definidos son de naturaleza numérica. La única condición especial ha consistido en que en el caso del organigrama el/la operador/a previamente debía ser entrenado en la interpretación de dichos gráficos, y en particular en la identificación del parentesco de los elementos que los constituyen. En este caso, a pesar de que se aseguró que previamente al registro de los datos los operadores eran capaces de identificar correctamente todas las relaciones familiares que posteriormente hallarían en las entrevistas, las deficiencias del registro sugieren que las interpretaciones "fueron libres" o difíciles debido a la baja calidad de los gráficos de que se disponía. En el primer caso, los resultados sugerirían que los operadores que capturan los datos nunca deberían interpretarlos otorgándoles un significado clínico-psicológico. En el segundo, que los entrevistadores que codificaron a mano la entrevista estaban utilizando un procedimiento no adecuado para registrar este tipo de información, o bien que carecían del suficiente entrenamiento.
3. En cuanto a las diferencias evidenciadas como consecuencia de la disponibilidad de distintos entrevistadores, estos resultados sugieren que cuando los datos se registran a mano la interpretación por parte del operador se dificulta. Parece lógico suponer, aunque no haya sido evaluado en este estudio, que la interpretación es más "libre" cuando se dispone de texto escrito a mano que cuando se dispone de texto mecanografiado o tratado con un procesador de textos. Esto nos ha sugerido que cuando la información se captura de fuentes documentales escritas a mano, sería práctico disponer de programas de captura que dispongan de un atributo especial para identificar los campos que han sido difíciles de leer/interpretar. En particular, este atributo debería diferenciarse del atributo de "valor provisional" y del código de "valor pendiente": (1) el atributo de valor "difícil de interpretar" sería un carácter especial al que recurrirían los operadores de captura de datos cuando tuvieran dificultades para interpretar un determinado valor -por ejemplo con objeto de consultar fuentes adicionales-; (2) el atributo de valor "provisional" sería un carácter de campo especial al que recurrirían los entrevistadores que utilizaran directamente el ordenador para la captura de datos -este atributo podría designar, tal como su nombre indica, que la información grabada en un determinado campo es provisional mientras no se disponga de nuevas evidencias sobre el valor correcto-; y (3) el código de valor "pendiente", que será comentado en el siguiente capítulo, también sería un código al que recurrirían los entrevistadores cuando capturan directamente los datos con ordenador -en concreto para señalar que la información es desconocida en el momento de grabar un registro, y que por lo tanto el campo queda "pendiente de ser grabado" tras la consulta de otras fuentes adicionales o de la misma fuente en otro momento-. La disponibilidad de estos atributos, en especial del primero, permitiría efectuar interesantes estudios sobre los errores de registro que son debidos a valores difíciles de interpretar, ya que durante la correspondiente captura los operadores los habrían marcado de forma especial. En su próxima versión comercial, el Sistema

DAT dispondrá de estos tres atributos/códigos; en su versión actual, ya dispone de un código de valor pendiente.

4. Finalmente, aunque en este estudio hemos obtenido diferencias en la proporción de discrepancias-errores según la sección de la entrevista (en particular, el apartado de demográficos ha sido el más difícil de registrar), consideramos que estos resultados no son concluyentes. La razón fundamental estriba en el hecho de que, precisamente, la sección de demográficos constituye el lugar donde se ubican las multirrespuestas, y en particular el organigrama. Por esta razón, es muy probable que el incremento evidenciado en la proporción de discordancias y errores esté confundiendo dos efectos bien distintos: el efecto de la sección con el efecto del tipo de campo.

En síntesis, los resultados de este trabajo apuntan que aunque la doble entrada y la verificación aleatoria tienen un coste de tiempo más elevado que el de la simple entrada, es importante incluir dichos sistemas de control (u otros similares) cuando se realiza la captura de datos en formato informatizado, ya que de este modo es posible evitar (aunque sea parcialmente) el efecto GIGO del que hablamos en el capítulo anterior. Esta afirmación se apoya, fundamentalmente, en los resultados obtenidos al valorar el segundo objetivo del estudio: los errores persistentes tras la verificación de los datos y su distribución. En este sentido, se ha comprobado que al comparar la Fase B verificada con el patrón la proporción de errores disminuye notablemente respecto a la comparación de la Fase A (sin verificar) con el patrón.

Pero no podemos finalizar este apartado sin realizar otra importante apreciación. La literatura sugiere que la doble entrada constituye la única estrategia que asegura que los datos estén libres de errores antes de proceder a su análisis (Blumenstein, 1993; Gassman et al., 1995; Gibson et al., 1995; Wolf, 1993). Los resultados de esta experiencia, sin embargo, constatan que esta estrategia no conlleva necesariamente que los datos introducidos coincidan plenamente con los datos registrados en el formato de papel. Así lo ha constatado la comparación de los datos verificados frente al correspondiente patrón: en todas las condiciones de registro persistían errores. En el siguiente capítulo se presenta una estrategia que demuestra ser válida para eliminar esta fuente de error: el uso de entrevistas informatizadas para la captura directa de los datos.

Antes de concluir este apartado también nos gustaría destacar otra de las deficiencias de gran parte de los sistemas de captura y gestión de datos disponibles a nivel comercial, y que podría incrementar notablemente la calidad de la gestión de datos clínicos: un registro de variables "histórico". Habitualmente, cuando los usuarios recurren a programas de captura de datos solicitan sistemas cuyo diseño facilite la grabación de la información, de forma que no sea necesario disponer de grandes conocimientos en informática para manejarlos de forma eficiente. Es cierto que la captura de la información tiene que automatizarse tanto como sea posible, pero también es absolutamente necesario de cara a la calidad de la gestión, la necesidad de registrar todas aquellas incidencias que se producen durante la entrada de un dato. En este sentido, sería ideal disponer de sistemas que, además de generar y ofrecer mensajes de aviso y/o error ante datos inconsistentes o no coincidentes entre dos entradas, también permitieran disponer de un archivo o campo de incidencias "histórico" en el que quedarán grabados dichos mensajes, el momento y la causa de su aparición, y las acciones correspondientes emprendidas por los usuarios para solucionarlos. Aunque la versión del Sistema DAT con la que hemos estado trabajando carecía de dicho registro histórico, la próxima versión comercial lo incluye.