

5 DISCUSIÓN

5.1 DISCUSIÓN SOBRE EL MÉTODO

La importancia del método de recopilación de datos en la elaboración de tablas y bases de datos de composición de alimentos ha sido demostrada y remarcada por diversos autores (8,9,15,16,79,140,228). En este campo, los trabajos publicados en España hasta la fecha se reducen únicamente a uno (191) que describe la estructura de una base de datos de composición de alimentos y que incluye elementos de interés metodológico. Por otro lado, no se ha publicado ningún trabajo que ponga a prueba a través de su aplicación las recomendaciones que diferentes autores y organismos internacionales (9,15,79) han ido realizando sobre la elaboración de tablas y bases de datos de composición de alimentos.

La colaboración entre compiladores y la capacidad para intercambiar datos entre diferentes bases de datos son fundamentales para optimizar el esfuerzo dedicado a la elaboración de BDCA ya que permiten, cuando puede ser de interés, compartir datos entre bases de datos (19,22,24,34). Sin embargo, utilizando la metáfora del lenguaje, la capacidad de intercambio de datos sólo es posible si las BDCA hablan el mismo lenguaje y utilizan el mismo vocabulario. Aquí radica la importancia de estandarizar estructura y códigos (27,28,35,79) utilizados en las BDCA (descripción de los alimentos, códigos descriptivos, modos de expresión) y de desarrollar interfaces que faciliten el intercambio de datos (124,136,187,229). No obstante, los estándares diseñados para la clasificación y descripción de los alimentos, descripción de los métodos analíticos, métodos de cálculo, identificación de los tipos de valores, procedimientos de estimación, factores de retención, etc. tienen que ser sometidos a prueba y refinados a través de su utilización (20,27,28,35) de manera que pueda llegarse a un acuerdo en estos puntos y a su aplicación a nivel nacional (230). En el presente trabajo se ha llevado a cabo la aplicación de algunos de estos estándares internacionales en el tratamiento de datos de composición y las observaciones extraídas se discuten en este apartado, así como algunos de los aspectos más relevantes del SGBD desarrollado.

5.1.1 Aplicación de los códigos EUROFOODS utilizados para describir los valores y los métodos

Para realizar la descripción de los valores (Tabla 3-8) y de los métodos usados para obtener estos valores (Tabla 3-2 y Anexo IV) se utilizan los códigos propuesto por EUROFOODS. Éstos códigos están diseñados para permitir un tratamiento ágil de la información durante el proceso de compilación. Sin embargo, como los propios autores señalan (30), han de ser

probados y refinados a través de su aplicación en sistemas como el descrito en el capítulo de Materiales y métodos. Se decidió incorporarlos dentro del SGBD teniendo en cuenta su carácter de recomendaciones de ámbito europeo pero también para identificar sus posibles puntos débiles. A continuación se detallan algunas de las observaciones recogidas de su aplicación.

5.1.1.1 Tipo de valor

En las recomendaciones EUROFOODS se define el tipo de valor (Tabla 3-8) como un código para describir el valor elegido como “mejor elección”. Sin embargo, los tipos de valores describen en realidad características diferentes de los valores. Se pueden distinguir tres grupos:

1. Los tipos de valores que hacen referencia al proceso de obtención del valor considerado la “mejor elección” por parte del compilador: MN (media), MD (mediana), MI (mínimo), MX (máximo), W (ponderación) y BE (mejor estimación).
2. Los tipos de valores que indican la situación de un valor: LT (menos que), MT (más que), TR (trazas), BL (por debajo del límite de detección), UD (sin resolver), N (desconocido).
3. Los tipos de valores que hacen referencia en realidad a estimaciones: LZ (cero lógico) y RZ (cero legal).
4. Los tipos de valores indeterminados: E (otros tipos de valores) y X (tipo de valor desconocido).

Los diferentes códigos no forman un conjunto de atributos que hagan referencia a una misma faceta de los valores, sino que en realidad hay tres grupos (el cuarto grupo no puede considerarse una categoría aparte). Esto podría dar lugar a situaciones en que a un mismo valor se le pudieran adjudicar dos atributos, creándose un conflicto. Durante el proceso de compilación se han producido situaciones en las cuales el código TR hubiera tenido que utilizarse simultáneamente con los tipos de valores MN, MD, MI, MX, W y BE. Esto ha llevado a considerar la posibilidad de que sea necesario utilizar simultáneamente los códigos MI, MX y BE y los códigos LT, MT y BL.

Por otro lado, es muy discutible incluir los códigos LZ y RZ dentro de los tipos de valores cuando en realidad tendrían que considerarse como tipos de métodos (Tabla 3-2) y, más concretamente, estimaciones con cierto riesgo de error. No se realiza una distinción entre

cero lógico y cero asumido, dos categorías de estimaciones diferentes. En realidad, esta distinción no se ha encontrado en la literatura consultada (9,15,30,81), aunque algunas fuentes la apuntan considerando el cero lógico como un caso específico del cero asumido (15).

Todas las estimaciones llevan asociadas cierta probabilidad de error, también las estimaciones en las que asignamos un valor cero a un componente. En algunas de ellas la probabilidad de error será cero o prácticamente cero (por ejemplo, cuando asignamos un valor cero de proteína al agua mineral embotellada), pero en otros casos la probabilidad de error puede no ser nula. Por lo tanto, el compilador “asume” que el valor verdadero es cero, pero también asume que existe un riesgo de error. La diferencia entre el valor verdadero y el valor estimado será el valor del error. Por otro lado, cuando asignamos un cero “lógico” (por ejemplo, cuando asignamos el valor cero a los ácidos grasos porque el valor de lípidos es cero) la estimación en sí no puede ser errónea ya que parte de una definición (“los ácidos grasos forman parte de los lípidos”) y de una proposición lógica que admitimos como cierta (“cuando los lípidos son cero, los ácidos grasos son también cero”). En todo caso, el valor del cual se parte puede ser erróneo y esto provocará que el cero lógico sea también erróneo. Todas estas consideraciones indican una dependencia entre estimaciones que han de tenerse en cuenta cuando se realizan o bien se revisan, y también la necesidad de diferenciar entre cero lógico y cero asumido (véase el apartado 3.9.7).

5.1.1.2 Tipo de método

Las recomendaciones EUROFOODS no dan ninguna definición para el tipo de método (Tabla 3-2) y se limitan a dar los códigos y describir cada uno de los tipos de métodos. Sin embargo, es fácil deducir que como tipo de método se quiere incluir cualquier procedimiento (analítico, cálculo o estimación) que se haya utilizado para generar un valor de composición. Naturalmente, si esta definición se acepta como la sugerida por las Recomendaciones EUROFOODS, se tendrían que incluir los ceros lógicos (LZ) y los ceros legales (RZ).

En la utilización de los tipos de métodos se han detectado algunas inconsistencias:

1. **Sumación de componentes que lo constituyen (S):** la definición de este tipo de método advierte que como sumación se incluye también la sustracción. Esto significa que los valores de carbohidratos obtenidos a partir de la suma de monosacáridos, disacáridos, oligosacáridos y polisacáridos recibirán el mismo código de tipo de método que el cálculo de carbohidratos por diferencia. El primero es considerado por las

recomendaciones EUROFOODS como una estimación del valor de carbohidratos a partir de los valores parciales de sus componentes, mientras que el segundo es una estimación que considera que son carbohidratos aquello que no es ni proteína, ni grasa, ni agua, ni fibra, ni cenizas. En algunos casos esto puede ser más o menos cierto, pero en otros casos no. El cálculo por suma de componentes es un estimación directa mientras que el cálculo por diferencia es una estimación indirecta en la cual el valor generado acumula los errores analíticos de los componentes usados en el cálculo (148). Sería recomendable disponer de códigos separados para cada uno de estos dos métodos.

2. **Calculado a partir del perfil de un componente (P):** este tipo de método es ambiguo en su aplicación ya que se utiliza tanto para cálculos realizados a partir de valores analíticos como para los realizados a partir de estimaciones. Por ejemplo, en el caso de los ácidos grasos puede referirse tanto a los valores calculados a partir de los ácidos grasos expresados respecto a 100 gramos del total de ácidos grasos procedentes de un análisis del mismo alimento, o bien procedentes de otro alimento que se utiliza como fuente de datos suponiendo que las diferencias en el perfil de ácidos grasos no son significativas (por ejemplo, utilizar el perfil de ácidos grasos de la leche entera pasteurizada para calcular el perfil de ácidos grasos de un queso fresco). En el primer caso se trata de una estimación que supone una cierta cantidad de ácidos grasos por gramo de grasa y en el segundo caso, además, que los ácidos grasos procedentes de otro alimento no difieren significativamente. Por otro lado, si se trata de agrupaciones de ácidos grasos, no queda claro si debe utilizarse el código P o el código S, todo depende en qué orden se realicen las operaciones (suma y conversión o al revés). Sería recomendable disponer de códigos diferenciados para cada caso que permitieran discriminar adecuadamente los diferentes métodos de obtención de los valores.
3. **Estimado de acuerdo con la legislación vigente (L):** tal y como están establecidas actualmente las recomendaciones EUROFOODS, este código tiene que utilizarse cuando el tipo de valor tiene un código RZ. Esto resulta en una duplicidad innecesaria de información.
4. **Estimado por deducción lógica (U):** tiene que utilizarse conjuntamente con el tipo de valor LZ, produciéndose una duplicidad de información innecesaria. Este problema se solucionaría incorporando el código LZ al grupo de códigos para el tipo de método.

5.1.1.3 Abreviatura del método

El diseño propuesto por las recomendaciones EUROFOODS (79) establece que cada valor de composición podrá tener un método asociado y que cada método podrá tener una abreviatura del método. Sin embargo esta estructura es insuficiente para describir adecuadamente los métodos analíticos y la utilización del código de abreviatura del método ha puesto de evidencia que para un método analítico puede ser necesaria la utilización de dos o más códigos de abreviatura del método.

Cada método analítico se puede descomponer en etapas analíticas que pueden ser variables. Por ejemplo, un método analítico para determinar ácidos grasos puede describirse con el código ME46 (CGL, columna capilar). Sin embargo, la extracción de la grasa previa al análisis puede efectuarse mediante el método Soxhlet (ME84) o bien el método de Folch (ME40). Como las dos combinaciones (ME84+ME46 y ME84+ME40) no son equivalentes y los resultados pueden variar según el método de extracción, para describir adecuadamente el método tendría que ser posible asociarle a cada valor de composición tantas abreviaturas del método como fueran necesarias. Actualmente, el SGBD sólo permite asociar un código de abreviatura del método pero puede introducirse texto libre que permite añadir la información que pueda faltar.

5.1.2 Códigos utilizados para describir los componentes

Los primeros códigos utilizados para describir los componentes en la BDCA fueron los códigos para componentes INFOODS. Se elaboraron con la intención de que sirvieran para facilitar el intercambio de datos entre BDCA. Pero antes de ser adoptados extensamente tenían que ser sometidos a prueba y mejorados a través de su utilización. Los códigos INFOODS están siendo utilizados en algunas bases de datos (17,48,140) desde hace unos años, pero ciertamente las experiencias son limitadas y existe muy poca revisión crítica de los códigos (79,160). Cuando se inició el diseño de la BDCA se decidió adoptarlos como códigos de identificación de los alimentos debido a su carácter de estándar internacional y a que ayudaban a sistematizar la identificación de los componentes dentro del SGBD.

No obstante, estos códigos revelaron un importante inconveniente a medida que eran utilizados. El origen de este inconveniente se halla en que el código INFOODS es una entidad débil que está conformada por hasta dos entidades diferentes de tres posibles: el componente, el método de obtención (analítico, estimación o cálculo) y la forma de expresión del valor. No todos los códigos tienen dos entidades: algunos sólo tienen una

(NIA, niacina) y otros dos (FATCE, grasa obtenida por el método Soxhlet; CHOAVLM, carbohidratos disponibles expresados como monosacáridos). La razón expuesta por los autores para diseñar el código de este modo es que se pretende diferenciar aquellos valores que no son equivalentes aunque se refieran al mismo componente (27). Así, por ejemplo, la grasa obtenida mediante el método Soxhlet tendría el código FATCE mientras que la grasa obtenida mediante hidrólisis ácida y el método Soxhlet tendría el código FAT.

Esta decisión presenta dos inconvenientes principales:

1. **Se repite información de manera innecesaria**, ya que el método de obtención o la forma de expresión se introducen también dentro de la base de datos en sus propios campos.
2. **Se dificulta la salida de datos**, ya que cuando se procede a agregar los valores hay que decirle de algún modo al sistema que FAT y FATCE deben ponerse en una misma pantalla. El compilador será quien decida si agrega o no los valores. Para conseguir esto es preciso establecer un nuevo código que sea el mismo para FAT y FATCE, pasando el código INFOODS a un segundo término.

El código para los componentes INFOODS parece más adecuado para la presentación de los valores cuando no se quiera incluir información sobre el método analítico o para un intercambio sencillo y limitado de datos. Cuando fue diseñado (década de los 80) las capacidades de los ordenadores y las posibilidades de intercambio de información eran mucho más limitadas. Actualmente las posibilidades de los actuales sistemas informáticos permiten almacenar mucha más información al mismo tiempo que permiten acceder fácilmente a ella, haciendo innecesario la reunión de varios códigos en uno sólo. De hecho, algunas aplicaciones recientemente diseñadas por otros autores dividen y separan el código de algunos componentes en dos códigos: uno para el componente y el otro para su forma de expresión (160). Además, los códigos para los componentes propuestos recientemente por las recomendaciones EUROFOODS (79) aplican el criterio de asignar un código diferente en función del método de obtención del valor en un caso muy especial: los carbohidratos calculados por diferencia (CHOT). En el resto de componentes, el código es independiente del método de obtención. Así, los valores de grasa tienen siempre el código FAT sea cual sea el método analítico utilizado (Soxhlet, Folch, etc.)

En el SGBD desarrollado se empezó utilizando los códigos INFOODS para identificar los nutrientes. Este código no presentó inconvenientes en una primera etapa ya que la salida de datos era sencilla y no se realizaban agregaciones. Sin embargo, cuando esta salida se

mejoró para permitir el cálculo de medias y ponderaciones, el código INFOODS presentó el inconveniente de que dos datos pertenecientes a un mismo nutriente pero con código INFOODS diferente no se podían situar en la misma plantilla para calcular la agregación (véase la Figura 3-21). Esto se solucionó adoptando el código EUROFOODS y utilizándolo para realizar la agregación de datos.

5.1.3 Utilización de formularios para la recopilación y primer tratamiento de los datos

La decisión de utilizar formularios de papel para realizar la primera recopilación y tratamiento de los datos fue motivada principalmente por la enorme dificultad que representaba diseñar y elaborar un SGBD que fuera capaz de realizar automáticamente la normalización de los valores de composición y a las limitaciones técnicas del sistema de gestión MS Access para incorporar la complejidad que tal tratamiento implicaba. Hay que tener en cuenta que los valores de composición pueden encontrarse expresados de maneras muy diferentes y que, aunque su normalización no requiere operaciones matemáticas complejas, sí es necesaria la participación de otros componentes y de factores de conversión específicos, que pueden variar según los alimentos o componentes y de los cuales existen diferentes versiones según los autores.

Los formularios permiten realizar la normalización manualmente de una manera metódica y controlada que permite revisar fácilmente los valores y operaciones. De este modo, los datos se introducen ya normalizados dentro del SGBD y esto hace innecesaria la incorporación de funciones de normalización dentro del SGBD. Sin embargo, este diseño tiene ciertas desventajas (Tabla 5-1) la principal de las cuales es el tiempo que requiere rellenar los formularios, realizar operaciones matemáticas muy a menudo repetitivas y revisar los formularios para evitar errores de transcripción y cálculo.

Por tal razón, en los casos de los azúcares y los ácidos grasos, se decidió introducir la posibilidad de realizar cálculos a través del SGBD. Se da la circunstancia adicional que en ambos casos la conversiones requerían la generación de nuevos componentes debido a la estructura del código de componentes INFOODS, lo cual hacía más tediosa su normalización manual ya que tenían que generarse nuevas filas dentro del formulario en lugar de utilizar las columnas “Datos sin normalizar” y “Datos normalizados” (Figura 3-3).

En definitiva, el trabajo con formularios permite salvar una dificultad técnica pero introduce un importante “cuello de botella” en el proceso de recopilación y tratamiento de los datos.

Tabla 5-1 *Ventajas y desventajas de la utilización de formularios para la recopilación y normalización de datos.*

VENTAJAS	DESVENTAJAS
Hace innecesario que el SGBD disponga de funciones de normalización, muy complejas de diseñar y elaborar.	Se tienen que realizar operaciones y cálculos repetitivos que podría realizar un sistema informático, ahorrando tiempo y errores.
Obliga al compilador a un examen exhaustivo y detallado de los datos.	Obliga a tener y mantener un voluminoso archivo de papel.
Permite una alta flexibilidad en la anotación de datos y observaciones en comparación a un sistema informático.	La actualización de datos o cálculos en los formularios puede ser imposible por el tiempo que implica, especialmente en el caso de actualizaciones masivas.
Fácil de proteger de posibles copias si lo comparamos con una base de datos informática.	Siempre es susceptible de pérdida de algún formulario o anotaciones. Los datos tienen que introducirse y revisarse dos veces, con lo cual se requiere más tiempo y se facilita la ocurrencia de errores de transcripción.

5.1.4 Sistema de evaluación de la calidad de los datos

El sistema de evaluación de los datos presenta ventajas importantes.

- 1. Introduce una sistemática en la evaluación de los datos que ayuda al compilador a centrarse en los aspectos más importantes.** El examen de una larga serie de datos que vienen acompañados de abundante información resulta muy difícil si no se dispone de un protocolo claro para seleccionar la información que realmente ayuda a valorar la adecuación de un dato.
- 2. Permite localizar fácilmente los datos que pueden merecer más atención durante la agregación de éstos (Figura 3-28).** De este modo pueden descartarse los datos con una calidad baja y no agregarlos con datos de una calidad superior. Evidentemente, el compilador tiene que fijarse también en la información detallada sobre cada dato antes de decidir qué valores y tipo de agregación son los más óptimos.
- 3. Permite auditar fácilmente la calidad global de la BDCA localizando aquellos grupos de alimentos y/o de componentes que presentan los códigos de calidad más bajos** y que son susceptibles de merecer mayor atención por parte del compilador. Esta posibilidad facilita mucho el mantenimiento y mejora de la BDCA.

Los inconvenientes del sistema de evaluación que se han ido poniendo de manifiesto a medida que se han ido utilizando son los siguientes:

1. **Se aplica el mismo sistema para todos los componentes sin tener en cuenta la especificidad de cada componente.** Aunque en un principio se creyó que esto no representaría un problema ya que los criterios que se utilizaban eran muy generales, posteriormente se comprobó que esto no era así. Por ejemplo, el número de muestras necesario para obtener un valor representativo puede venir determinado en parte por las características del componente. Así, por ejemplo, componentes lábiles como la vitamina C pueden requerir un mayor número de muestras para la obtención de valores representativos (70,71).
2. **Se aplica el mismo sistema a todos los alimentos.** La aplicación de los mismos criterios de muestreo y número de muestras se realiza asumiendo que todos los alimentos son más o menos iguales por lo que se refiere a producción y distribución, pero es obvio que esto no es así (9,70). En el caso de alimentos producidos por una o dos empresas que tienen la mayor parte de la cuota de mercado se intentó compensar esta tendencia a asignar puntuaciones bajas introduciendo la nota (a) en la Tabla 3-3.
3. **El peso de los diferentes criterios en la puntuación final es el mismo.** Aún no existe en la literatura un método o un sistema que permita determinar cual debe ser el peso que cada criterio ha de tener. Aunque se trata de una variable categórica, la puntuación de calidad se utiliza como si fuera numérica, sumando las puntuaciones parciales y esto obliga a establecer ponderaciones. Una posible solución sería precisamente desarrollar un sistema de evaluación de la calidad de los datos que no requiera realizar estas ponderaciones. Esto se conseguiría renunciando a un único código de calidad y conservando la información de cada aspecto (Tabla 3-3) por separado, opción que permitiría almacenar más detalles sobre cada dato.

5.2 DISCUSIÓN SOBRE LOS RESULTADOS

El núcleo principal de los resultados del presente trabajo es la BDR generada por cuanto contiene todos los datos y metadatos. No se trata de la primera BDCA realizada en España: existe noticia de que se elaboró una BDCA española a principios de la década de los 90 utilizando el modelo relacional y el sistema de gestión de base de datos DB-III (191). Esta BDCA incluía datos de composición de alimentos y metadatos (descripción del alimento, componente, método analítico y fuente de los datos). Sin embargo, este trabajo no se concretó en una BDCA destinada a los usuarios habituales de datos de composición, ni se tiene constancia de que finalmente se elaborara una TCA o se utilizara en algún estudio nutricional. Parece ser que esta BDCA se utilizó únicamente para uso interno o bien no se continuó su desarrollo.

La BDCA desarrollada es la primera destinada a la población española que se ha elaborado utilizando un método sistemático de evaluación de la calidad de los datos analíticos y que incorpora recomendaciones internacionales en su estructura y contenido, hecho que hace posible el intercambio de datos con otras bases de datos que sigan tales recomendaciones. Además, se ha generado una base de datos del usuario con la composición de 698 alimentos para 35 componentes, sin valores desconocidos y documentada. Esta BDU puede incluirse directamente en programas informáticos de análisis de dietas. A partir de la BDU se ha elaborado una TCA impresa con toda la información necesaria para su correcta utilización.

5.2.1 Recopilación de documentos

El examen minucioso de las fuentes de información permite descartar de entrada algunos datos que pueden introducir sesgos en los valores de composición o cuya documentación no es suficiente para poder realizar agregaciones de manera conveniente (Tabla 4-1). El porcentaje de documentos en los que se daba este caso es considerable (un 11% de los documentos tratados) y las razones pueden ser diversas, pero las relacionadas con una deficiente documentación son las que podrían remediarse más fácilmente. El resto o bien dependen de los objetivos particulares de la publicación o bien se trata de errores.

El problema de la documentación reaparece cuando se examinan los datos individualmente y no la publicación en su conjunto. Algunas fuentes han sido aceptadas y sus datos incorporados al SI, pero la documentación de ciertos aspectos ha resultado incompleta. El

caso más importante es el del método analítico, en el cual hasta un 13% de los datos han recibido un código de tipo de método “desconocido” (X) debido a que la fuente no proporcionaba información detallada. Por ejemplo, sólo se suministraba una referencia bibliográfica o referencias vagas (por ejemplo, “método oficial” o “normas ISO”). En otros casos se suministraba una referencia a un método oficial o a una referencia bibliográfica. Buscar la referencia, conseguirla y extraer la información sobre el método analítico implicaba mucho tiempo y sólo se realizó cuando se podía localizar fácilmente la referencia.

Otro caso particular de documentación incompleta concierne a los alimentos con marca. Mientras que algunos autores, los menos, ofrecen datos acerca de las marcas analizadas o incluso la marca comercial, otros no ofrecen tal información. La información sobre la marca comercial es muy útil para realizar las agregaciones y asegurar la representatividad de los valores. Cuando esta información no se ofrecía en la publicación se preguntó a los autores sobre las marcas comerciales pero no se obtuvo respuesta en ningún caso.

Se han encontrado otros problemas en la recopilación de datos que han impedido la utilización de algunos de ellos también relacionados con documentación incompleta o incorrecta (Tablas 4-1 y 4-2). Por ejemplo, cuando se proporcionan datos referidos a materia seca sin proporcionar el contenido en agua (haciendo imposible convertir los datos referidos a materia fresca), la falta de coherencia de los valores (por ejemplo, la suma de los nutrientes mayoritarios es mayor/menor de 100 ± 3) o la ambigüedad en las formas de expresión de algunos datos (utilización del término “porcentaje” o “%” en lugar de las unidades del Sistema Internacional).

Para evitar los problemas generados por las causas citadas en la recopilación de datos destinados a BDCA pero también en la utilización de datos por otros autores, es preciso que tanto autores como revisores trabajen para mejorar la documentación que se ofrece en los artículos científicos que contienen datos de composición (71,126,231).

Las empresas alimentarias se consideraron al principio como posibles buenas fuentes de información. Sin embargo, se detectaron algunos inconvenientes cuando se les solicitó información sobre sus productos:

1. **Las empresas suelen no atender las peticiones de datos sobre la composición nutricional de sus productos.** De un total de 167 empresas se obtuvieron datos de 26 de ellas (16%). Las causas son básicamente ausencia de datos, razones de confidencialidad (aunque los datos que se solicitan sean los mismos que los que figuran

en la etiqueta nutricional de los productos) o bien la imposibilidad de preparar la documentación solicitada debido al día a día. Es importante remarcar que cuando se solicitaban los datos de composición también se solicitaban los metadatos necesarios para evaluarlos. Proporcionar toda esta información suponía, en la gran mayoría de casos, un esfuerzo importante por parte de la empresa o sencillamente no disponían de ellos.

2. **La cantidad de información sobre la composición de sus productos parece ser muy limitada.** Excepto algunas de las grandes compañías, la mayoría de empresas disponen de información muy básica.
3. **Cuando disponen de datos analíticos, a menudo los métodos utilizados son inadecuados u obsoletos.** Esto suele ocurrir porque los laboratorios donde se encargan los análisis tienen métodos de rutina establecidos por regulaciones legales que no son adecuados en el ámbito nutricional o bien son métodos anticuados pero más económicos (por ejemplo, el método Carr-Price para determinar retinol).
4. **Es frecuente que los datos de composición que disponen las empresas se hayan obtenido a través de datos procedentes de TCA u otras fuentes, pero no son datos analíticos.** Esto ocurrió en el 27% de las empresas que enviaron datos de composición. El Real Decreto que establece como debe elaborarse el etiquetado nutricional señala que *“las cifras declaradas deberán ser los valores medios obtenidos, según el caso, a partir de: a) El análisis del alimentos efectuado por el fabricante; b) El cálculo efectuado a partir de valores medios conocidos o efectivos de los ingredientes utilizados; c) Los cálculos a partir de datos generalmente establecidos y aceptados”* (211). Por esta razón se decidió que todos los datos procedentes de etiquetas se tratarían como estimaciones y no como datos analíticos (véase el apartado 3.9.10).

Actualmente, estos inconvenientes ponen en entredicho la utilización de los datos procedentes de empresas alimentarias como fuente masiva y fiable de información. Y esto representa un serio problema por cuanto el consumo de alimentos manufacturados en Europa es cada vez mayor, estimándose en la actualidad que puede situarse en un 43-63% del total de ingesta de alimentos (230) y que la diversidad de alimentos en los mercados occidentales también se está incrementando (9). El alto coste que representa obtener muestras y analizarlas hará que resulte cada vez más difícil renunciar a la información procedente de las empresas por lo que parece conveniente que estas participen de alguna manera en la elaboración de BDCA o bien que se organicen para elaborar bases de datos

sectoriales que puedan ser usadas en el etiquetado nutricional, tal y como se está intentando en EEUU (62,232,233).

Los artículos científicos como fuente de información tienen la ventaja de ser fáciles de conseguir, y ofrecer bastante información sobre los datos de composición (Tablas 4-11 a 4-13). No obstante, a menudo la información ofrecida por los autores es insuficiente para una adecuada evaluación de los datos (126), lo cual puede resultar en que no sea posible evaluar los datos y queden excluidos (véase el apartado 4.1) o que la puntuación que reciban sea menor a la que realmente podrían recibir (véase el apartado 4.2.3). Cuando se requiere alguna información sobre los datos directamente de los autores no siempre se obtiene una respuesta y, además, tales recopilaciones de información adicional consumen mucho tiempo (95). Las tesis aportan también muy buena información y, a veces, más detallada que las publicaciones correspondientes, pero son difíciles de conseguir. Los datos de laboratorios son muy difíciles de conseguir, por razones obvias de confidencialidad, y la información que se da sobre los valores (y que se utilizaría como metadatos) es muy poca.

5.2.2 Base de datos de referencia

El método indirecto utilizado en la elaboración de la BDCA permite establecer cierta prioridad en la recopilación de datos de composición de alimentos ya que normalmente la búsqueda de información existente se realiza por alimento. Sin embargo, es difícil establecer prioridades en cuanto a componentes y normalmente se tienen que introducir y tratar todos los componentes considerados por la fuente de información. El SGBD desarrollado permite, no obstante, registrar todos los componentes incluidos en la fuente e introducir posteriormente sólo aquellos que interesan a corto término. Sin embargo, se observó que esto no representaba un ahorro de tiempo importante ya que donde realmente se dedicaba tiempo era en la introducción de datos en el formulario y, en este caso, era mejor no dejar de lado los componentes no prioritarios ya que esto implicaba que, tarde o temprano, se tendría que efectuar una nueva lectura de la fuente de información. Además, llevar un control de los componentes no introducidos era bastante más complicado y engorroso que con el SGBD.

En el caso de los ácidos grasos, la disparidad entre fuentes de información en cuanto a los isómeros analizados era considerable, aparte del hecho de que se trata del grupo más conflictivo en cuanto a nomenclatura de los componentes debido precisamente al número y diversidad de isómeros (234). Es muy frecuente que la nomenclatura se utilice incorrectamente y se comentan errores como el de denominar "ácido oleico" al grupo de isómeros con fórmula C18:1 (octadecenoico). Esta disparidad en la utilización de la

nomenclatura introduce una indeterminación importante que obliga a realizar comprobaciones con los datos e incluso a descartarlos para su utilización. Aunque finalmente se hayan recopilado valores hasta para 118 isómeros y agrupaciones de isómeros, las diferencias en las agrupaciones de isómeros y los problemas de nomenclatura complican extremadamente la realización de agregaciones y de estimaciones. Por tal razón se decidió trabajar de momento con las agrupaciones clásicas en saturados, monoinsaturados y poliinsaturados.

La base de datos permite la creación de históricos completamente documentados sencillamente copiando las tablas principales de la base de datos. Tales históricos son útiles para realizar estudios retrospectivos teniendo en cuenta las posibles variaciones en la composición de los alimentos y la mejora en los datos de composición, para lo cual es imprescindible la documentación asociada a los datos de composición (235).

Aunque se ha intentado concentrar la recopilación de datos en aquellos grupos de alimentos más importantes, el método de elaboración sólo utiliza datos ya existentes. Así puede observarse que en algunos de los grupos de alimentos importantes se han conseguido más datos de origen español que en otros (Tablas 4-8) lo cual repercute en el código de calidad de los datos (Tabla 4-15). Los grupos para los cuales existe menos información sobre alimentos españoles publicada y recopilada hasta la fecha son "Cereales y derivados", "Tubérculos y derivados" y, en menor grado, "Verduras y hortalizas" y "Frutas y derivados". Sin embargo, esto también está correlacionado con el hecho de que las fuentes de información que se utilizan cuando no se disponen de datos españoles son bases de datos y tablas de composición extranjeras, muchas de las cuales no ofrecen información sobre el método analítico.

Estas consideraciones pueden extenderse a los componentes, en los cuales los grupos de vitaminas ofrecen las puntuaciones de calidad más bajas. En este caso destacan las vitaminas liposolubles y las vitaminas hidrosolubles como grupos para los cuales existe muy poca información publicada y referida a alimentos españoles que haya sido introducida en la BDR hasta la fecha (Tabla 4-9), la mayor parte de ella procedente de tablas de composición españolas (aproximadamente un 80% de todos los datos españoles publicados sobre vitaminas e introducidos hasta la fecha). Esta limitada disponibilidad de datos sobre vitaminas se corresponde con una falta de datos publicados, aunque es difícil cuantificar su importancia exacta hasta que no se tengan más datos introducidos en el sistema.

5.2.3 Base de datos del usuario y tabla de composición

La cantidad de datos procedentes de fuentes españolas de la BDR es del 33,1% (Tabla 4-7). En el caso de la BDU esta cifra es más difícil de estimar ya que los valores se agregaban y en tal agregación podían participar valores correspondientes a alimentos españoles o no. Se realizó una aproximación tal y como se indica en el apartado 4.3.1. y el resultado fue que un 32,1% de los valores de la BDU que no son estimaciones son valores correspondientes a alimentos españoles. Si lo referimos al total de alimentos de la BDU (incluyendo los valores estimados), el valor se sitúa en un 24,8%.

Pocas bases de datos informan con detalle acerca del origen de los datos. La tabla de composición EPIC para España (236), elaborada principalmente a partir de otras tablas de composición y con algunos datos procedentes de artículos y otras fuentes, indica que un 15,8% de los datos proceden de TCA o artículos españoles. Evidentemente, algunas de estas fuentes utilizan a su vez datos procedentes de TCA extranjeras. Por otro lado, el banco de datos italiano desarrollado para su uso en epidemiología (16) incluye un 45,3% de datos procedentes de tablas o artículos italianos y contiene unos 900 alimentos. De nuevo, las tablas utilizadas como fuente principal (sólo un 0,02% de los datos proceden de artículos) incluyen a su vez datos procedentes de TCA extranjeras. La BDCA neozelandesa contiene unos 2500 alimentos y el porcentaje de datos procedentes de fuentes neozelandesas es de un 42%, de fuentes extranjeras un 33% y estimados un 25% (88). No se dispone de más información acerca de otras tablas o bases de datos.

El porcentaje de ceros directos (código LZ, denominados ceros lógicos en la codificación de EUROFOODS; véase el apartado 3.9.7) es bastante alto (15,4%) e indica la importancia de realizar este tipo de estimación para evitar la presencia de valores desconocidos que pueden ser fácilmente suprimidos. Así, por ejemplo, algunas TCA no los incluyen provocando la aparición de valores desconocidos de manera absolutamente innecesaria (144,200,237,238). En algunos casos, la cantidad de valores desconocidos que podrían ser estimados como ceros directos es realmente alta (200,237).

La posibilidad de agregar los datos para calcular el valor final que se incluirá en la BDU permite incrementar la representatividad de los valores. Partiendo siempre de datos validados, cuantas más fuentes se utilicen y mayor número de muestras analizadas, mayor representatividad de los valores agregados (15). La cantidad de datos resultantes de una agregación (tipo de valor MN, AV o W) se situaba en un 8,6% de los valores de la BDU. Si contamos además con los datos que proceden de una sola fuente pero que se han obtenido analizando más de una muestra, la cantidad se incrementa a un 13,4%. Este grupo de

valores serían los más representativos dentro de la BDU. Evidentemente, en la representatividad de los valores también influye el criterio seguido para establecer las agregaciones. Sin embargo, en la mayoría de casos no se detectaron valores fuera de rango que pusieran en duda la representatividad de la agregación, aunque sí se daban algunos casos en que, como ya se ha comentado (véase el apartado 4.2.2) la deficiente descripción de las muestras impedía realizar ponderaciones que hubieran mejorado aún más la representatividad de los valores agregados.

Los valores con tipo de método desconocido ("X") son numerosos (46,2%) debido a la importancia de las TCA como fuente de datos. Esto limita de manera importante el escrutinio de los datos respecto al método analítico, el metadato más importante cuando se trata de establecer la validez de un dato, y obliga al compilador de datos a basarse en el prestigio del organismo que elabora la TCA. Una de las prioridades más urgentes sería, por lo tanto, conseguir disminuir este porcentaje a niveles muy inferiores y a ser posible limitarlos a alimentos con poco peso en la dieta española.

Uno de los aspectos más destacables de la BDU es la ausencia total de valores desconocidos. Esto la convierte en una base de datos óptima para su uso en programas de cálculo de dietas ya que permite calcular la ingesta de nutrientes evitando el riesgo de infraestimación debido a la presencia de valores desconocidos (14,16,51,239-241). Algunos estudios que han evaluado el efecto de la sustitución de los valores desconocidos por el valor cero en BDCA de uso habitual han encontrado infraestimaciones que se sitúan en el rango del 0,04% al 14% según el componente (242) o incluso llegar a más del 25% en el caso de la fibra alimentaria (243). No obstante, la importancia de esta infraestimación a menudo no es fácil de detectar por el usuario a partir de los datos finales calculados.

En un estudio comparativo de 6 programas de análisis de dietas (3 norteamericanos, 1 francés y 2 españoles) se pudo comprobar como la presencia de valores desconocidos puede llevar a la obtención de valores con un sesgo importante que comporta la infraestimación de la ingesta de algunos nutrientes (244). En este estudio se calcularon los aportes nutricionales de tres dietas de un día con diferentes aportes energéticos (2.000, 3.500 y 4.000 kcal) usando los 6 programas mencionados. En la Tabla 5-2 se han incluido los valores obtenidos con los dos programas españoles: "Nutrición" (245) utiliza las tablas de composición de Mataix y colaboradores mientras que "PCN 1.0" (246) utiliza la BDU elaborada por el autor.

Si se comparan los valores obtenidos pueden observarse la existencia de discrepancias importantes en los resultados obtenidos. Sin embargo, es difícil apreciar cuales son debidas

a los valores desconocidos y cuales a otras razones como, por ejemplo, la variabilidad de la composición de los alimentos, la utilización de diferentes fuentes de datos, la inclusión de datos posiblemente obsoletos, los errores en las tablas, etc. (198,218). Si nos limitamos a los nutrientes para los cuales el programa “Nutrición” presenta un porcentaje de valores desconocidos superior a 30%, podremos observar comportamientos bastante diferentes: nutrientes en los que claramente se aprecia una infraestimación en los cálculos elaborados con “Nutrición” (ácidos grasos saturados, ácidos grasos monoinsaturados, colesterol, vitamina E) y otros en los que no parece producirse una infraestimación o incluso parece que el programa que en realidad genera valores más bajos es “PCN” (ácidos grasos poliinsaturados, vitamina D, niacina y cinc).

Para esclarecer estas aparentes contradicciones es preciso examinar a fondo los datos. Si se calcula el cociente entre el valor total de lípidos y la suma de ácidos grasos se observa que mientras el resultado obtenido para “Nutrición” varía entre 0,67 y 0,79, en el caso de PCN se mantiene constante en 0,92-0,93. Este cociente indica la proporción de ácidos grasos en los lípidos de la dieta. Como nuestro aporte principal de lípidos son las grasas de adición y las animales intrínsecas (Tabla 5-3), el valor debería situarse cerca de 0,95. Un valor cercano a 0,7 no es coherente ya que solo podría conseguirse si la fuente principal de grasas fuera el pescado blanco, determinadas vísceras animales y la harina de cereales (9). Puede concluirse que en los resultados obtenidos con el programa “Nutrición” todos los ácidos grasos están claramente infraestimados.

En el caso de la niacina hay que tener en cuenta que los datos del programa “Nutrición” incluyen la posible niacina procedente de la bioconversión del triptófano. Por lo tanto, tendrían que ser sistemáticamente más altos que los datos de “PCN”, pero en la dieta de 4000 kcal incluso es inferior a los datos de “PCN”. En el resto de componentes la situación es variable y pueden haber diferencias que se mantienen constantes como en el caso de la vitamina E, evidenciando también una posible infraestimación (la cantidad de valores desconocidos en este caso es del 40%), hasta diferencias que aparecen y desaparecen o incluso se invierten como en el caso de la vitamina D y el cinc. Estos últimos comportamientos pueden ser debidos a la elección de los alimentos: en una dieta se eligen alimentos para los cuales hay pocos o ningún valor desconocido, mientras que en otra dieta existe mayoría de valores desconocidos. A esto hay que añadir que los datos de los alimentos no procesados (verduras, carnes, pescados, etc.) de la BDCA del programa “Nutrición” corresponden en su práctica totalidad a alimentos crudos, mientras que en el programa “PCN” se incluyen alimentos crudos y cocinados. Evidentemente, este hecho afectará la estimación de la ingesta de vitaminas y minerales produciendo un sesgo cuya

Discusión

dirección es difícil de predecir. Es importante que las BDU y TCA incorporen alimentos en las formas que son consumidos (51,141,247).

Tabla 5-2 Comparación de los resultados obtenidos del análisis de 3 dietas diferentes (corresponden a 1 día) utilizando dos programas informáticos: "Nutrición" y "PCN". El primero utiliza la TCA de Mataix y colaboradores que contiene un porcentaje considerable de valores desconocidos en determinados nutrientes, mientras que el segundo utiliza la BDCA elaborada por el autor y que no contiene valores desconocidos (244). La composición de cada dieta puede encontrarse en la Tabla 5-3.

Componentes	Unidades	2.000 kcal		3.500 kcal		4.000 kcal		Valores desconocidos (%)
		Nutrición	PCN	Nutrición	PCN	Nutrición	PCN	Nutrición
Energía	kcal	2010	2115	3615	3445	4075	3900	0
Agua	g	- (3)	1670	-	2370	-	2740	-
Carbohidratos (1)	g	301	293 (316)	533	471 (485)	616	529 (568)	2
Azúcares	g	-	122	-	220	-	270	-
Polisacáridos	g	-	171	-	231	-	259	-
Proteína bruta	g	103	107	174	169	164	176	1
Proteína vegetal	g	-	38	-	66	-	71	-
Proteína animal	g	-	69	-	103	-	105	-
Lípidos	g	51	56	102	107	123	120	2
AGS	g	10,5	18,8	20,6	29,5	19,8	31,8	36
AGM	g	17,9	25,2	32,6	39,2	41,5	47,7	40
AGP	g	5,60	8,35	27,2	29,8	29,6	31,7	40
Cociente Σ AG : Lípidos	-	0,67	0,93	0,79	0,92	0,74	0,93	-
Colesterol	mg	172	222	254	351	234	355	47
Fibra	g	35	37	48	52	54	58	25
Vitamina A	μ g e.r.	1920	1846	1158	1172	1353	1421	18
Retinoides totales	μ g e.r.	-	216	-	260	-	294	-
Carotenoides totales	μ g e.b.	-	9775	-	5465	-	6763	-
Vitamina D	μ g	1,0	1,3	2,0	1,4	0,4	2,4	36
Vitamina E	mg	6,22	8,41	11,4	12,2	13,7	15,7	40
Tiamina	mg	1,7	2,3	3,0	2,4	2,7	3,3	23
Riboflavina	mg	2,4	2,3	3,4	3,4	3,1	4,1	14
Niacina (2)	mg	37	25	57	37	47	48	32
Vitamina B6	mg	4,0	3,3	5,4	4,4	4,6	5,2	29
Ácido fólico	μ g	520	620	480	750	450	920	23
Vitamina B12	μ g	3,4	3,1	6,9	6,1	4,7	6,2	20
Vitamina C	mg	310	260	290	240	300	270	24
Sodio	mg	1490	1690	2060	3120	2570	3420	14
Calcio	mg	1170	1110	1700	1690	1580	1530	9
Magnesio	mg	372	400	787	627	775	636	23
Potasio	mg	4230	4490	6600	5830	6770	6590	18
Hierro	mg	18,8	18,5	32,2	25,35	31,4	28,6	19
Cinc	mg	7,54	9,7	20,1	17,3	17,8	18,2	35

(1) Los carbohidratos en "Nutrición" están expresados en equivalentes de monosacáridos, mientras que en "PCN" están expresados como tales. Entre paréntesis se incluye el resultado de la suma de los azúcares multiplicados por 1,05 y los polisacáridos multiplicados por 1,1; el valor es una buena aproximación a los carbohidratos expresados en equivalentes de monosacáridos; (2) Los datos de "Nutrición" incluyen la posible niacina procedente del triptófano, mientras que en "PCN" no se contabiliza; (3) Ausencia de datos. Abreviaciones: AGS, ácidos grasos saturados; AGM, ácidos grasos monoinsaturados; AGP, ácidos grasos poliinsaturados; e.r., equivalentes de retinol; e.b., equivalentes de betacaroteno.

Si se deseara comprobar el efecto de la presencia de valores desconocidos sobre las estimaciones de ingesta sería necesario trabajar con largas series de datos que eliminaran el efecto de la elección de los alimentos. Concluyendo, el usuario puede no percibir el efecto de la presencia de valores desconocidos en sus cálculos aunque éste exista y, por consiguiente, es importante garantizarle que la base de datos no contiene valores desconocidos o, como mínimo, que su porcentaje es bajo en todos los componentes y que se limitan a alimentos de consumo menos frecuente. Siempre es preferible una estimación a la presencia de un valor desconocido que los programas informáticos traducirán por un valor cero (3,14).

Tabla 5-3 Lista de alimentos y cantidades (g) que componen las tres dietas analizadas en la Tabla 5-2. Las cantidades se refieren a porción comestible de alimento.

2.000 kcal		3.500 kcal		4.000 kcal	
Alimento	g	Alimento	g	Alimento	g
Aceite de oliva	20	Aceite de oliva	30	Aceite de oliva	40
Arroz blanco	60	Azúcar	30	Almidón de maíz	40
Azúcar	20	Carne magra de ternera	150	Almidón de maíz	30
Copos de maíz tostados	30	Cebolla	25	Azúcar	30
Judía verde	250	Copos de maíz tostados	40	Cebolla	25
Leche entera UHT	240	Espaguetis	100	Jamón cocido	40
Lechuga	100	Leche desnatada UHT	240	Leche UHT entera	500
Lenguado	150	Leche entera UHT	240	Lechuga	125
Lenteja	40	Lechuga	125	Lenguado	180
Manzana Golden	100	Lenguado	180	Lenteja	80
Naranja	100	Lenteja	75	Macarrones	100
Pan blanco	120	Maíz	50	Maíz	50
Patata vapor	125	Mermelada de ciruela	30	Manzana Golden	250
Plátano	100	Miel	25	Mermelada de ciruela	30
Pollo asado	100	Naranja	125	Miel	25
Queso de bola	30	Nueces sin cáscara	45	Naranja	125
Tomate crudo	100	Pan blanco	230	Nueces sin cáscara	45
Yogur desnatado	125	Patata al vapor	125	Pan blanco	240
Zanahoria	100	Patata al vapor	150	Patata al vapor	150
Zumo de naranja comercial	200	Pera	125	Pera	125
		Plátano	125	Plátano	125
		Queso de bola	30	Queso de Burgos	30
		Rábano	25	Rábano	25
		Tomate crudo	100	Ternera estofada	150
		Yogur natural	250	Tomate crudo	100
		Zanahoria	50	Yogur natural	250
		Zumo de naranja comercial	200	Zanahoria	50
				Zumo de melocotón	240
				Zumo de naranja comercial	200

Respecto a los componentes principales incluidos finalmente en la BDU y las tablas de composición su adecuación a las necesidades actuales que establecen los diferentes usos establecidos como prioritarios en las especificaciones (véase el apartado 3.11.2.1) puede determinarse comparando la lista de componentes incluidos en la BDU y las TCA con las listas de componentes incluidos en las recomendaciones nutricionales, documentos que

establecen objetivos nutricionales o normas legislativas realizadas por diferentes organismos (Tabla 5-4).

La BDU y la TCA cubren todos los componentes incluidos en las recomendaciones españolas (207) excepto el yodo. En el caso de las recomendaciones europeas (206) los componentes no incluidos son el yodo, ácidos grasos de la serie n-3 y ácidos grasos de la serie n-6, biotina, ácido pantoténico, selenio, cobre, manganeso. Respecto a las recomendaciones internacionales de la FAO/WHO (157,203-205), se excluyen el yodo, ácidos grasos de la serie n-3, ácidos grasos de la serie n-6, biotina, selenio, ácido pantoténico, vitamina K, linoleico y alfa-linolénico.

Respecto a las recomendaciones establecidas por sociedades científicas españolas, objetivos nutricionales y normas legislativas (59,208-211), los componentes no incluidos en la BDU y las TCA son el yodo, el flúor y los ácidos grasos de la serie n-3 seguidos de los ácidos grasos de la serie n-6, la biotina y el cobre.

Aunque estos nutrientes no se han podido incluir en la TCA si se han recopilado datos de composición para ellos (véase el Anexo II). Sin embargo, la cantidad de datos ha sido insuficiente como para poder incluirlos en el proceso de generación de la BDU y la TCA. La recopilación y tratamiento de datos para estos nutrientes permitirá ir incorporándolos en el futuro.

Tabla 5-4 Componentes incluidos en diferentes documentos que establecen recomendaciones u objetivos nutricionales y en legislación de etiquetado. Los componentes sombreados son los que se incluyen en la BDU y las tablas impresas. En la última fila se indica el porcentaje de componentes que cubren la BDU y la TCA. O: el componente se incluye en el documento/legislación; (O): el componente es necesario para el cálculo de otros componentes o para calcular proporciones, aunque no está incluido en las recomendaciones consultadas.

Componente	FAO WHO	RN (UE)	RN (ES)	SENC	DECV	ORCV	PS (CAT)	RD390	MSC
Energía	O	O	O	(O)	(O)	O	(O)	O	O
Agua									
Proteína bruta	O	O	O	(O)	O	O		O	O
Proteína animal									
Proteína vegetal									
Lípidos	O			O	O	O	O	O	O
AGS	O			O	O		O	O	O
AGM	O			O			O	O	O
AGP	O	(O)		O	O		O	O	O
AG n-3	O	O		O			O		
AG n-6	O	O		O					
Ácido linoleico	(O)								
Ácido alfa-linolénico	(O)								
Colesterol	O			O	O				O
Carbohidratos	(O)	(O)	(O)	O	(O)	O	(O)	O	O
Azúcares					O		O	O	O
Polisacáridos					O		O	O	
Polialcoholes	(O)	(O)	(O)	(O)	(O)	(O)	(O)	O	(O)
Ácidos orgánicos	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)	(O)
Fibra total				O	O		O	O	O
Etanol	(O)	(O)	(O)	(O)	O	O	(O)	(O)	O
Sodio				O	O		O	O	O
Potasio		O							O
Calcio	O	O	O	O		O	O	O	O
Magnesio	O	O					O	O	O
Fósforo		O					O	O	O
Hierro	O	O	O			O	O	O	O
Cinc	O	O	O			O	O	O	O
Cobre		O					O		
Manganeso		O							
Selenio	O	O							O
Yodo	O	O	O	O	O	O	O	O	O
Flúor				O			O		
Vitamina A	O	O	O			O	O	O	O
Retinoides									(O)
Carotenoides							O		(O)
Vitamina D	O	O	O			O	O	O	O
Vitamina E	O	O	O			O	O	O	O
Vitamina K	O						O		
Tiamina	O	O	O			O	O	O	O
Riboflavina	O	O	O			O	O	O	O
Niacina	O	O	O			O	O	O	O
Ácido pantoténico	O	O						O	
Vitamina B6	O	O	O			O	O	O	O
Ácido fólico	O	O	O	O		O	O	O	O
Vitamina B12	O	O	O			O	O	O	O
Vitamina C	O	O	O			O	O	O	O
Biotina	O	O						O	
Cobertura (%)	74	74	90	74	87	90	83	91	94

FAO/WHO: Recomendaciones realizadas por FAO/WHO (157,203-205); RN (UE): Recomendaciones de la UE (206); RN (ES): Recomendaciones españolas (207); SENC: Objetivos nutricionales para la población española de la SENC (59); DECV: Dieta y enfermedades cardiovasculares. Recomendaciones de la SEA (208); ORCV: Obesidad y riesgo cardiovascular. Recomendaciones de SEA-SEEDO-SEEN-SEMFYC-SEMI-SENC (209); PS(CAT): Plan de Salud de Catalunya (210); RD930: Real decreto 930/1991. Norma de etiquetado sobre propiedades nutritivas de los productos alimenticios (211). MSC: Directrices para la elaboración de estudios poblacionales de alimentación y nutrición (39).

5.2.4 Resultados claves del estudio

Los resultados claves de este estudio son los siguientes:

1. Se ha puesto a punto un Sistema de Información completo que permite generar bases de datos de composición de alimentos al mismo tiempo que:

- Controla la trazabilidad de los datos y su calidad, lo cual facilita el mantenimiento de la BDCA
- Permite su mejora y actualización continua, ya que las diferentes funcionalidades están pensadas para la incorporación continua de información al sistema
- Permite evaluar el estado de la base de datos, ya que toda la información se almacena en formato electrónico haciendo posible la elaboración de consultas complejas para extraer información sobre los datos (calidad, compleción de la BDCA, antigüedad de los datos, etc.)
- Permite el intercambio de datos con otras bases de datos, gracias a la incorporación de códigos estandarizados internacionalmente.
- Permite conservar históricos de la base de datos, ya que las diferentes versiones elaboradas pueden almacenarse fácilmente.

2. La aplicación de algunas recomendaciones internacionales (EUROFOODS e INFOODS) para el manejo e intercambio de BDCA ha permitido detectar problemas de diseño en los códigos.

3. Se ha conseguido generar una BDU y una TCA sin valores desconocidos, con datos para 35 componentes principales y 698 alimentos, con las siguientes características relevantes según los criterios establecidos por Greenfield y Southgate (3):

- La selección de valores efectuada durante el proceso de datos permite proporcionar datos de calidad analítica válida y representativos. Sin embargo, esto depende también de la posibilidad de acceder a fuentes de datos bien documentadas ya que la TCA se ha elaborado con el método indirecto. Actualmente este aspecto es mejorable si se disminuye el porcentaje de datos procedentes de tablas extranjeras.

- La lista de alimentos se ha elaborado utilizando como referencia los alimentos registrados en diferentes encuestas de consumo en España y abarca los alimentos de mayor consumo.
- La lista de nutrientes cubre la mayor parte de los nutrientes incluidos en las principales recomendaciones y objetivos nutricionales aplicables a la población española.
- La descripción de los alimentos se basa en el sistema por facetas de INFOODS para poder incluir toda la información que asegure una buena descripción de los alimentos.
- La expresión de los datos se define claramente para evitar ambigüedades.
- El origen de los datos, así como las operaciones realizadas sobre ellos se proporcionan a nivel de valores de composición en el caso de la BDU.
- En la TCA y BDU se proporciona información auxiliar para facilitar su uso.
- La utilización de estándares internacionales permite el intercambio de datos con otras BDCA que utilicen estos mismos estándares.
- La TCA y la BDU no presentan ningún valor desconocido.

5.2.5 Limitaciones del estudio

Las principales limitaciones del presente estudio son las siguientes:

- 1. La introducción manual de datos, su normalización y el cálculo de nuevos componentes en los formularios es tediosa y lenta.** Además, la actualización masiva de datos es prácticamente imposible.
- 2. La inclusión de datos obtenidos de alimentos españoles y la documentación de los valores (sobre todo en relación al método de obtención) son relativamente bajas.** Aunque los porcentajes de datos procedentes de alimentos propios se sitúan en valores similares a los de otras TCA elaboradas con el método indirecto, es evidente que

es preciso mejorar este aspecto. También es preciso disminuir el porcentaje de datos cuyo método de obtención es desconocido. Esto se conseguirá recopilando más datos analíticos originales documentados. Sin embargo, el hecho de utilizar datos existentes impone una limitación obvia.

- 3. No se pudo incluir la fuente de datos para cada componente en la TCA.** A ser posible, las fuentes de información de cada valor se tendrían que especificar para cada uno de ellos. En el caso de la BDU esto se pudo hacer sin demasiadas dificultades, aunque es preciso realizar una consulta (con Access o SQL) para obtener una relación de fuentes utilizadas para cada valor. Sin embargo, en el caso de la TCA esto comporta ciertas dificultades ya que se trata de compaginar el formato considerado más útil para el usuario (apaisado con disposición tabular de la información) con la inclusión de una lista de las referencias bibliográficas consultadas junto a cada valor, teniendo en cuenta que esta lista puede ser de hasta 10-12 referencias diferentes. Finalmente se decidió primar el formato por encima de la inclusión de las fuentes para cada valor ya que se consideró que esta información tendría un interés muy limitado para los usuarios principales de estas TCA.

5.2.6 Implicaciones de este trabajo

Cabría distinguir dos aspectos diferenciados en cuanto a las implicaciones que puede tener el presente trabajo. Por un lado, una tarea larga y compleja como es la recopilación y tratamiento de datos genera una cantidad importante de conocimiento que puede aprovecharse para mejorar el Sistema de Información o bien para el diseño de nuevos proyectos relacionados:

- A partir de las observaciones realizadas de la aplicación de las recomendaciones EUROFOODS se pueden proponer cambios y modificaciones en algunos aspectos de tales recomendaciones (véase el apartado 5.1.1).
- Una vez se hayan introducido la mayor parte de la información publicada sobre composición de alimentos será posible evaluar de manera precisa el estado del conocimiento sobre composición de los alimentos españoles. Esta información puede servir para determinar qué áreas tienen las carencias más importantes de conocimiento. Además, a partir de los valores recopilados es posible realizar una estimación de la variabilidad de la composición de alimentos (70). Ambas informaciones son necesarias

para planificar convenientemente la realización de un programa de muestreo y análisis para conseguir nuevos datos de composición a través del método directo (71,177).

Por otro lado, la BDU y la TCA que se han generado han sido utilizadas en las siguientes actividades:

- Elaboración de la publicación titulada “Tablas de composición de alimentos CESNID” (227), la cual es la primera TCA publicada en España sin valores desconocidos. En otros países se dispone de tales tablas orientadas específicamente a la evaluación de la ingesta de nutrientes de individuos o poblaciones (14,16) en las que los valores desconocidos han sido totalmente substituidos por valores procedentes de publicaciones o bien generados a partir de diferentes métodos de estimación. La publicación mencionada incorpora también un programa informático de cálculo que permite utilizar fácilmente los datos de las tablas, y tablas y bases de datos sobre porciones y gramajes de los alimentos que facilitan la realización de estimaciones de consumo de alimentos y su posterior conversión a ingesta de nutrientes.
- La BDU esta siendo utilizada en la elaboración de una base de datos de composición nutricional para el estudio EPIC (230). El estudio EPIC (*European Prospective Investigation into Cancer*) es un estudio de cohorte multicéntrico en el que participan 10 países europeos (Alemania, Dinamarca, España, Francia, Grecia, Holanda, Inglaterra, Italia, Noruega y Suecia) que tiene como objetivo estudiar la asociación causal entre diferentes localizaciones tumorales y el consumo de alimentos, bebidas alcohólicas, macronutrientes, micronutrientes y otros componentes de la dieta habitual (248-251). Esta participación ha sido posible gracias a la exhaustiva información recogida en la base de datos y a la incorporación de las recomendaciones EUROFOODS en su diseño, lo cual ha permitido el intercambio de datos con el resto de países participantes en el proyecto (213).

5.2.7 Sugerencias para posteriores estudios

A partir de las experiencias recopiladas a lo largo del estudio y de los resultados obtenidos es posible plantear mejoras y ampliaciones del Sistema Informático creado:

1. La única solución a la limitación impuesta por la utilización de formularios para introducir y tratar datos dentro del SI es **que el SGBD permita introducir los datos en bruto y que el mismo sistema se encargue de realizar los cálculos que sean necesarios**

para normalizar los datos o generar nuevos componentes, encargándose también de actualizarlos cuando sea necesario. Sin embargo se trata de una solución muy compleja ya que sería preciso que el SGBD fuera capaz de saber cuáles son las expresiones de los valores preferentes, cuáles son los factores de cálculo a utilizar y cómo se deben utilizar y que fuera capaz de detectar los cambios en los factores o en los datos que exigirían su re-normalización. Por sí solo ya constituye un proyecto de gran envergadura.

2. Existe un alto número de alimentos para los cuales no se pueden encontrar datos de composición y tampoco saldría a cuenta, si algún día fuera posible, su muestreo y análisis. **Si tales alimentos son complejos y se pueden desglosar en alimentos básicos para los cuales existe información de composición, es posible calcular la composición del alimento problema utilizando algoritmos de cálculo que tienen en cuenta los principales cambios nutricionales** que se producen durante su elaboración (15,77). Tal sistema tiene una amplia utilización en otras TCA (48,141,145) ya que permite estimar de manera bastante fiable la composición de muchos alimentos. Por otro lado, también hay que tener en cuenta que normalmente este tipo de alimentos tienen una composición en ingredientes y formas de preparación muy variables (por ejemplo, la tortilla de patatas o la paella de arroz) y que en términos de coste/beneficio es más conveniente calcular su composición que intentar realizar un muestreo representativo y analizar todas las muestras (4,15,77,141). De nuevo, diseñar e implementar tal opción en el SGBD es altamente complejo, más aún si se quiere que esté integrada en un sistema que permita realizar diversas operaciones con los datos y al mismo tiempo se desea que se mantenga la integridad y coherencia en ellos.
3. Del mismo modo que tendría que ser posible automatizar los cálculos de normalización y los de recetas, **tendría que ser posible automatizar determinadas estimaciones**. Por ejemplo, la asignación de valores cero sigue unos criterios que o bien son lógicos o son muy fáciles de definir (véase el apartado 3.9). El compilador solo tendría que definir los criterios y los alimentos destinatarios de un valor determinado y el mismo sistema se podría encargar de adjudicar dichos valores y de controlarlos. Teniendo en cuenta hasta un 15% de los valores pueden ser ceros directos (Tabla 4-16) la utilidad de tal aplicación no es despreciable.
4. Así como se ha incluido un sistema para evaluar los datos que se consideran de origen analítico, **tal vez sería preciso incluir también un sistema de evaluación de las estimaciones ya que no todas las estimaciones tienen el mismo valor de incertidumbre**. Por ejemplo, una estimación de tipo cero directo lógico no es

equivalente a una estimación realizada con un alimento parecido pero que no es la misma especie biológica. Del mismo modo que es interesante controlar la calidad de los datos analíticos, tendría que serlo controlar la calidad de las estimaciones. No existe constancia hasta la fecha de que se haya realizado un trabajo en este sentido.

5. Aunque se ha tenido una especial atención en establecer pautas para que los alimentos estén adecuadamente descritos, **sería preciso incorporar un sistema estandarizado de clasificación o descripción de los alimentos** (30,35,79,100). De los sistemas de descripción de alimentos existentes, el sistema LanguaL es especialmente interesante por su clara orientación al uso en sistemas informáticos y por su vocación internacional. No obstante, la implantación de este sistema dentro del SGBD no ha podido ser abordado en esta versión.
6. **Es necesario mejorar la descripción de los métodos analíticos con la finalidad de facilitar su validación.** El método analítico es el aspecto más importante cuando se evalúa la validez de un valor y, por lo tanto, es una información principal para la toma de decisiones del compilador. Sería preciso detallar la descripción procurando esquematizarla al mismo tiempo para facilitar su lectura e interpretación. Para ello sería útil considerar las diferentes fases que pueden constituir un método analítico (189,252) y describir cada fase utilizando texto libre al cual se haría corresponder un código unificado como el utilizado en la presente versión del SGBD (véase el Anexo IV) y una referencia bibliográfica. Se podría entonces permitir tanto la validación de los datos a través de la validación del código unificado como de la referencia bibliográfica o de ambos simultáneamente, desencadenando un proceso en cascada que validaría los datos dependientes. Evidentemente, pueden existir conflictos que se tendrían que resolver profundizando en el análisis. Un sistema así facilitaría enormemente la evaluación de los datos ya que la validación del método analítico no se realizaría dato por dato sino método por método.
7. **Deberían incorporarse los componentes yodo, ácidos grasos de la serie n-3, ácidos grasos de la serie n-6 y selenio** para mejorar la cobertura de las necesidades de los potenciales usuarios.

