

5.3. Die Datenbank

Mit der Eingabe und Kontrolle der erhobenen Daten hat man sich eine Datenbank geschaffen und die Daten können nun ausgewertet werden. Für große Studien ist dies allein schon ein zeitraubender, jedoch unumgänglicher Vorgang, der pro Studienteilnehmer viele Stunden Arbeit bedeutet.

Doch es müssen noch mehr Aufgaben bewältigt werden, damit die Datenbank funktionsfähig wird. Man benötigt noch entsprechende Auswertungsprogramme; die richtige Wahl hängt wiederum von der Datenstruktur ab, die schon bei der Dateneingabe bekannt sein muss, da davon auch die Art der Eingabe bzw. der Aufbau des Kode-Planes abhängt. Die entsprechenden Programme sind in Rechenzentren verfügbar bzw. sind käuflich zu erwerben. Für viele Fälle reichen diese Rechenprogramme, doch es bedarf in anderen Fällen weiterer Hilfe durch EDV-Experten.

Bei den meisten ernährungsepidemiologischen Untersuchungen ist die Daten-Struktur nicht von ideal rechteckiger Matrix, sondern es gibt eine Vielzahl von verschiedenen Variablen und Daten, die unterschiedlichen Gruppierungen von Studienteilnehmern und verschiedenen Erhebungszeitpunkten zugeordnet sind; nicht alle Teilnehmer nehmen an allen Erhebungseinheiten teil. Es existieren manchmal auch verschiedene Daten-Ebenen (Individuum/ Gruppe, Haushalt; oder: Lebensmittel/Rezept, Mahlzeit). Für solche heterogenen Daten-Mengen gibt es angepasste EDV-Datenverwaltungsprogramme, wie z. B. SIR (*Scientific Information Retrieval*).

Neben den Auswertungs- und Datenbank-Programmen sind noch weitere EDV-Programm-Systeme notwendig bzw. hilfreich. Man benötigt Schreib- bzw. Editier-Programme (wie XEDIT, WORD, usw.) und Graphik-Programme (wie Erlanger-Graphik-System, Harvard Presentation Graphik, usw.) zur Eingabe, Beschreibung und Darstellung von Daten und Informationen. Da in der Regel nicht alle Datenverarbeitungsanforderungen durch käufliche Software zu erfüllen sind, sind auch Programmier-Programme (wie FORTRAN, BASIC, usw.) notwendig. Schließlich braucht man noch Programme, die den EVD-Geräten Anweisungen geben, wie sie zu arbeiten haben, wie sie die andere Software in ihre "Maschinen-Sprache" übernehmen können (Betriebs-Programme, wie DOS, NOS, UNIX usw.).

Eine Datenbank benötigt weiterhin zum wirkungsvollen Auswerten von ernährungsepidemiologischen Studien eine Reihe von Vergleichsdaten, wie Referenz- bzw. Standardwerte zur Bewertung der einzelnen Variablen. Zur Bewertung der Nährstoff-Aufnahmen gibt es Empfehlungen für die Nährstoff-Zufuhr (z. B. RDA = recommended dietary allowances). Für die Körpermaße gibt es eine Reihe von anthropometrische Standards. Für die biochemischen Indikatoren gibt es ganze Tabellenwerke mit den entsprechenden Normwerten. Weiterhin sollten in den Datenbanken spezielle Informationen verfügbar sein, wie z. B. Dokumentation von Ereignissen (z. B. für den Aufbau von lokalen Kalendern zur Kontrolle von Alters- bzw. Zeitangaben) und Umweltdaten (z. B. Klimadaten zur Kenntlichmachung von saisonalen Einflüssen).

Im Prinzip ist es wünschenswert, für jede erhobene Variable Vergleichsdaten zu haben. Das entspricht auch dem allgemeinen Forschungsprozess und der Bildung des Untersuchungsmodells; vom Stand der Erkenntnis ausgehend, die Studie zu planen (Abb.15). Solche Vergleichsdaten sind Sekundärdaten (s. S. 94f). Es gibt hier einige Ansätze, neben den herkömmlichen Literatur-Sammlungen solche Informations-Fakten auf EDV-Datenbanken zu speichern und abzurufen.

Bei longitudinalen Studien entstehen solche Fakten-Datenbanken quasi automatisch, müssen doch die Daten aus den verschiedenen Erhebungs-Zeiträumen miteinander vergleichbar gemacht werden. Ein spezieller Bereich von Fakten-Datenbanken im Rahmen ernährungsepidemiologischer Studien sind - bedingt durch ihre Informationsfülle - solche für Ernährungsdaten.

Die Konzentrationen der vielen möglichen Lebensmittelinhaltstoffe sind für ernährungsepidemiologische Studien wichtige Informationen. Die Vielzahl der verschiedenen Lebensmittel einerseits und die Zahl der verschiedenen Stoffe für sich allein deutet schon den Umfang an Arbeit und Mühe an, der erforderlich ist, um Nährwert-Tabellen zusammenzustellen. Lebensmittel sind keine konstanten homogenen, sondern variable, biologische Materialien. Nur wenige Lebensmittel sind homogen und stabil - wie z. B. Kristallzucker. Selbst relativ homogen erscheinendes zeigt schnell Unterschiede; bei der Milch kann sich oben der Rahm absetzen, bei Obstsäften unten die Pulpe. Obst hat Schalen und Kerne; Brot hat Rinde und Poren, usw. Fleischstücke können offensichtlich sehr unterschiedlich sein, mageres kann mit Fett durchwachsen sein, man kann Knochen und Bindegewebe sehen. Noch heterogener sind die vielen Zubereitungen, man denke nur an verschiedene Eintöpfe oder an "Pizza mit Allem". So ist das erste wichtige Problem bei der Erstellung von Nährwert-Tabellen, dass man repräsentative Proben für die Analysen zieht. Dazu braucht man eine Reihe von Informationen, die ansich wieder eigene Vorstudien darstellen (Abb.55). Für jedes Lebensmittel gibt es eigene Überlegungen zur Stichproben-Ziehung - dies gehört zu den Aufgaben des Lebensmittelchemikers bzw. der Lebensmittel-Kontrolle.

Zu der Stichproben-Ziehung bzw. der notwendigen Dokumentation gilt es den Namen des Lebensmittels bzw. der Zubereitung eindeutig zu identifizieren (Abb.55). Das erscheint relativ einfach, doch selbst hier gibt es Probleme. Sicher am geringsten auf der Ebene der Nahrungsrohstoffe, da hier die verschiedenen möglichen lokalen bzw. einheimischen Namen mit Hilfe der internationalen biologischen Nomenklatur (Taxonomie) zu überprüfen sind. Aber bereits die Definition der verschiedenen essbaren Teile von Lebensmitteln kann in einzelnen Ländern unterschiedlich vereinbart sein. Es leuchtet ein, bei Zubereitungen als namentliche Definition nicht einfach nur "Weichkäse mit bestimmter Fettgehalt-Stufe" anzugeben, selbst Camembert, Brie, usw. sind zu allgemein. Analog sind Nudeln und "Spätzle" genauer zu definieren, wie auch die jeweilige Brot- oder Wurstsorte. Der gleiche Name kann in verschiedenen Regionen mit ganz anderen Rezepturen und Verarbeitungen verbunden sein.

So ist es notwendig, die einzelnen Bestandteile der Produkte bzw. Zubereitungen zu kennen, sowie möglichst detaillierte Angaben über die Herkunft der Lebensmittelproben zu haben, wobei man praktisch den Weg der Nahrungskette so nachvollziehen sollte, dass Inhaltsstoffbestimmende Vorgänge identifiziert werden können. Hier treten bei den heutigen internationalen Verflechtungen auf den Lebensmittel-Weltmärkten und -Produktionsstätten große Informationsprobleme auf. Schließlich muss man wissen, welchen Zustand die Lebensmittelprobe bei der Probennahme hatte.

Als nächster großer Informationsbereich müssen Angaben über die Art und Qualität der einzelnen Lebensmittel-Analysenschritte in die Überlegungen mit eingehen. Das betrifft den Weg der Probe in der Lagerungs-, Vorbereitungs- und der eigentlichen Analysenphase. Auch hier müssen einige Andeutungen genügen, um den Problembereich aufzuzeigen. So kann der Energiegehalt einerseits experimentell ermittelt werden (kalorimetrische Messung), er kann aber auch aus den Brennwerten (z. B. ATWATER-Faktoren) für die einzelnen energieliefernden Nahrungsinhaltsstoffe berechnet sein. Der Gehalt an Kohlenhydraten kann

analysiert sein, oder aus der Differenz (Eiweiß, Fett, Asche zu Trockengewicht) berechnet sein. Die Kohlenhydrat-Analysen sind für die jeweiligen Kohlenhydrat-Gruppen (Mono-, Oligo-, Polysaccharide; Zucker, Stärke, Rohfaser, usw.) getrennt zu betrachten. Für jeden (Nähr)Stoff gibt es verschiedene Methoden, die Qualität der einzelnen Methoden ist unterschiedlich; je nach Entwicklung der Analysen-Methoden sind die ermittelten Werte von Zeit zu Zeit zu revidieren. In letzten Jahren betraf dies besonders die Beurteilung der Werte für Rohfaser und Folsäure.

Schließlich müssen noch Angaben über die Verzehrsgewohnheiten für die Stichprobenziehung berücksichtigt werden; es sollen schließlich essbare biologische Materialien untersucht werden. Es gibt bekanntlich international gesehen viele essbare Lebensmittel. Es können nicht alle für alle (Nähr-)Stoffe untersucht werden, man muss Prioritäten setzen und die bestimmen sich einfach durch das, was am häufigsten verzehrt wird. Grundnahrungsmittel haben höhere Priorität gegenüber denen, die nur ganz selten bzw. in geringen Mengen verzehrt werden. Das ist aber keine konstante und homogene Einteilung (andere Länder - andere Sitten - andere Verzehrsgewohnheiten). International gesehen ist die Spannbreite von Alltagsspeisen groß. Die Speisepläne ändern sich mit der Zeit, es kommen nicht nur immer wieder neue phantasievolle Lebensmittel-Verarbeitungen auf den Markt, wie verschiedene Produkte der "weißen Regal-Linie" (Joghurts, Quarks, Milchcremes, usw.) oder der "Müsli-Bio-Welle", sondern Exotisches kommt auch direkt aus fernen Ländern (Kiwi, Litschi, usw.). Die modernen Völkerbewegungen (Tourismus, Gastarbeiter, Flüchtlinge, Asylanten, usw.) bringen fremde Ernährungsgewohnheiten mit und so ergeben sich entsprechende Aufgaben für die Erstellung von Nährwert-Tabellen.

Aber nicht nur die durchschnittliche Verzehrmenge allein soll als Information herangezogen werden, sondern auch individuelle Verzehrsmengen, d.h. z. B. welche Portionsgrößen werden im allgemeinen von welchen Personengruppen (Alter, Geschlecht, Lebensstil, usw.) verzehrt. Bei statistisch geringen Verzehrsmengen pro Kopf und Tag kann in einzelnen Fällen doch die individuelle Verzehrsmenge relativ groß sein. Einige Stichpunkte sollen diese Aussage verdeutlichen: die Bestandteile von Festtagsessen (z. B. Gans, Truthahn zu Weihnachten), von besonderen Altersgruppen (z. B. Muttermilch bei Säuglingen) und die Besonderheiten in den Ernährungsgewohnheiten von Anhängern spezieller Kostformen, wie z. B. Vegetariern, Makrobioten, usw. Als weitere Beurteilungsebene kommt noch hinzu, dass man nicht nur allein die verzehrten Mengen absolut betrachten sollte, sondern auch den Gesichtspunkt des Beitrages an der jeweiligen Nährstoffversorgung. Verschiedene Lebensmittel haben verschiedene Nährstoffdichten. So ist bei Obst der Vitamin C Gehalt von größerer Bedeutung als der Eiweißgehalt oder der Energiebrennwert.

All diese vorgenannten Informationen dienen dazu, den nötigen Umfang der Stichproben für die Lebensmittel-Analyse abzuschätzen, bzw. später die Repräsentativität des Wertes in den Nährwert-Tabellen abschätzen zu können. Das Prinzip der Stichprobengrößen-Berechnung ist das gleiche, wie bei der Ziehung der Stichprobe für die Studienteilnehmer (s.Kap.2.3.3.). Je größer die Variabilität der Produkte, desto mehr Proben sind nötig; im Einzelnen ist dazu jedoch ein weiter Weg von Überlegungen anzustellen; dies ist wiederum nur mit Hilfe der entsprechenden Experten möglich.

Es muss schließlich überlegt werden, in welcher Weise die einzelnen Analysenwerte gespeichert werden. Die verschiedenen Analysen der einzelnen Probe eines gleichen bzw. sehr ähnlichen Lebensmittels bzw. Zubereitung sollten zusammengefasst werden, wobei Mittelwert und Variationsbreiten angegeben werden. Es ist zu überlegen, wie "tief" die Nährwerttabellen bzw. Datenbanken zu differenzieren sind. Auf der Ebene der EDV-Speicher

kann dies recht weit nach unten geschehen, d.h. die Werte sind nach der Herkunft, Verarbeitungsstufe, usw. sortiert; die Speicherkapazität lässt dies zu. Bei den gedruckten Tabellen bzw. bei der Berechnung der erhobenen Ernährungsdaten ist dies nicht angebracht. Je nach Studiensituation und Studienziel kann die Aggregation aller verschiedenen Apfelsorten zu „Apfel“ genügen, wenn es z. B. um die Nahrungsenergieaufnahme geht. Steht jedoch der Vitamin C Gehalt im Mittelpunkt, dann sind, bedingt durch den unterschiedlichen Gehalt an Vitamin C in den einzelnen Apfelsorten, entsprechende Untergliederungen nötig.

Betrachtet man den Umfang der vielen verschiedenen Informationen über Lebensmittel und ihre Inhaltsstoffe, bzw. die notwendigen begleitenden Randinformationen (Abb.55), dann kommen heute für die Archivierung solcher Daten nur noch EDV-Systeme in Frage. Die "festen" Labor-Werte müssen mit einem "Kranz" von Informationen umgeben werden, damit man die "Fakten" beurteilen und einordnen kann.

Die Analysen der einzelnen Lebensmittel ergeben nicht stabile Größen; die Analysen sind teilweise schwierig; unterschiedliche Methoden ergeben unterschiedliche Werte. Die Lebensmittel sind nicht konstant. So gibt es unterschiedlich gute Nährwert-Angaben zu den einzelnen Lebensmitteln (*Stewart* 1983). Verschiedene Tabellen-Werke kommen auch zu verschiedenen Angaben, weil sie auf andere Labordaten zurückgreifen, andere Zuordnungen vornehmen, kurz die Informationen über die Stichproben anders bewerten (z. B. *Rand et al.* 1987).

Für ernährungsepidemiologische Studien sind Ernährungs-Datenbanken zur Berechnung der Nährstoffaufnahmen notwendig. Aus diesem Grunde hat sich praktisch jede entsprechende Forschergruppe aus den vielen vorhandenen Nährstoff-Tabellen-Werken und eigenen Nährstoff-Analysen bzw. -Sammlungen individuell eigene Datenbanken aufgebaut, die unterschiedlichen Aufbau und Umfang haben (z. B. EUROFOODS 1985). Das ist zwar kein prinzipieller Fehler, doch es hat einige Nachteile. Das Sammeln entsprechender Informationen und das Installieren dieser Daten in eine entsprechende Datenbank ist eine sehr aufwendige und nie zu Ende gehende Arbeit. Das wurde in diesem Kapitel beschrieben. Eine Zusammenarbeit in diesem Bereich würde unnötige Wiederholungen ähnlicher Tätigkeiten vermeiden und darüber hinaus sind solche Absprachen nötig, um die Ergebnisse der Studien verschiedener Untersuchungsgruppen vergleichbar zu machen. Dieselben Ernährungserhebungsdaten mit unterschiedlicher Verschlüsselung und Nährwert-Tabellen berechnet, ergeben rein rechnerisch unterschiedliche Ergebnisse. Allein schon die Benutzung ähnlich aufgebauter EDV-Nährwert-Berechnungsprogramme führt zu merklichen Unterschieden.

Es sind weltweit Bestrebungen im Gange, die Ernährungs-Datenbanken zu standardisieren. Da die Anwender bzw. die Nutzer jedoch ganz verschiedene Studiensituationen und ganz unterschiedliche Studienziele haben, müssen die gemeinsam genutzten Ernährungs-Fakten-Datenbanken sehr differenziert aufgebaut sein, damit die einzelnen Informationen von den jeweiligen Nutzern entsprechend ihren Bedürfnissen zusammengestellt werden können. Nicht nur die Datenbank-Inhalte sondern auch die Verschlüsselung müssen dann abgestimmt werden, wobei die wünschenswerte international-koordinierten und -gültigen Übereinkünfte auf einige wesentliche Problembereiche stoßen, die sich vor allem aus der verschiedentlich genannten Vielzahl der Lebensmittel bzw. der Unterschiede in den Ernährungsgewohnheiten ergeben. So sind je nach Verzehrsgewohnheiten und Untersuchungsziel unterschiedlich detaillierte Zuordnungen angezeigt. Deutlich wird das am Beispiel der Kartoffel, die in Deutschland ein Grundnahrungsmittel darstellt; sie kommt in sehr vielen verschiedenen Sorten und Zubereitungen vor und kann so eine eigene Lebensmittelgruppe in den

Nährwerttabellen bzw. im Nahrungsmittelschlüssel darstellen. In anderen Ländern in denen sie wenig verzehrt wird, scheint es angebracht, sie anderen Wurzelgemüsen zuzuordnen. Folglich müssen internationale Vereinbarungen sehr detaillierte Informationen speichern (Truswell 1988).

Aus solchen umfassend aufgebauten Datenbanken, die quasi als "Referenz-Datenbanken" dienen, müssen sich relativ leicht (z. B. durch entsprechende Zusammenfassungen der Detailintragungen; höhere Stufe in der Schlüssel-Hierarchie) die jeweils individuell notwendigen "Arbeits-Datenbanken" ableiten lassen. So sind für manche Fragestellungen, wie z. B. bei der einfachen Abschätzung der Nahrungsenergie-Aufnahme nur eine Kurzform von Nährwerttabellen nötig. In solchen Fällen mögen die Angaben für 100 Lebensmittelgruppen über die Nahrungsenergie-Gehalte genügen. Bei internationalen Verbundstudien können dagegen lange Listen aus verschiedenen nationalen Datenquellen die Arbeitsgrundlage sein. Schließlich wird es in anderen Fällen notwendig sein, sich nicht mit den in den Datenbanken gespeicherten Informationen zu begnügen, sondern man muss eigene Nährwert-Analysen mit in die Studie bzw. in die "Arbeits-Datenbank" aufnehmen (z. B. "Total Diet Studies", s.S.184).

Entschließt man sich in ernährungsepidemiologischen Studien zu eigenen Nährwert-Analysen, dann muss hier natürlich auch die Qualität der eigenen Analysen und gespeicherten Informationen kritisch gewürdigt werden. Das fängt bei den Überlegungen zur Probeziehung an. Man muss sich überlegen, welche Mahlzeiten in welcher Form (einzelne Lebensmittel, roh, "Einkaufskorb"; fertige Gerichte, ganze Tagesportionen) zur Analyse vorbereitet werden. Auch bei der chemischen Analyse der Nahrungsprouben zeigen sich bemerkenswert unterschiedliche Ergebnisse. Werden beispielsweise aliquote Teile einer Mahlzeit analysiert, so müssen deshalb im Haushalt größere Mengen zubereitet werden als üblicherweise verzehrt werden ("*duplicate portion*" für die Analyse). So kann die untersuchte Probe u.U. nicht der ursprünglichen Zusammensetzung der Speisen entsprechen. Die Wahrscheinlichkeit dafür wächst, je heterogener eine Mahlzeit zusammengesetzt ist. Die Gefahr einer fehlerhaften Probenahme ist dann besonders groß, wenn diese von den Studienteilnehmern selbst vorgenommen wird. Außerdem muss man berücksichtigen, welche Teile der Zubereitungen mit welcher Wahrscheinlichkeit nicht mitverzehrt werden (Fett, Knorpel, usw.). Selbst bei guten Probenahmen ergeben sich zwischen gleichen Duplikaten Abweichungen, die allein aus der Inhomogenität der Lebensmittel entstehen (Schlage et al. 1988). Weitere Fehlermöglichkeiten ergeben sich dadurch, dass die Lebensmittelproben durch Umwelteinflüsse oder durch Mineralstoffe, die sich aus (Koch)Gerätschaften lösen und dann als Inhaltsstoffmengen den Lebensmitteln zugeschlagen werden, verfälscht werden. Diese Gefahr ist z. B. bei Eisen gegeben.

Die Lebensmittel-Analysen sind recht schwierig, da es sich wie bereits mehrfach erwähnt, bei Lebensmitteln um komplexe Stoffgemische handelt. Auf mögliche Störungen durch andere Inhaltsstoffe muss bei den vielen möglichen Lebensmitteln jeweils individuell geachtet werden, der Stand der Analysetechnik ändert sich. Auch daraus erklären sich Unterschiede in einzelnen Nährwerttabellen bzw. Ernährungs-Datenbanken.

Die Ernährungs-Datenbanken müssen verschiedenen Ansprüchen gerecht werden, die nicht ohne weiteres miteinander vereinbar sind. So ist einerseits die Vergleichbarkeit, d.h. Stabilität der gespeicherten Informationen wichtig. Nur so können Studien an verschiedenen Orten (internationale Zusammenarbeit) und zu verschiedenen Zeiten (longitudinale Studien) ohne methodischen Fehler miteinander verglichen werden. Andererseits müssen die Datenbanken die "natürlichen" Veränderungen bei den Lebensmitteln (z. B. Angebotsänderungen) bzw. in

der Analysetechnik schnell mit aufnehmen, d.h. aber Flexibilität. Dieser Zielkonflikt - Stabilität vs. Flexibilität - lässt sich dadurch lösen, dass die Sachinformationen in der Datenbank mit den zeitlichen Dimensionen in Beziehung gebracht werden.

Insgesamt ist es notwendig, dass der Inhalt von Ernährungsfakten-Datenbanken, der bisher hauptsächlich auf Inhaltstoff-Angaben beschränkt ist, mit weiteren Informationen ergänzt wird, die für die Beurteilung der Daten (z. B. zur Probenermittlung, Abb.55) und Ergebnisse von Bedeutung ist; dies betrifft insbesondere die Verzehrsmengen der verschiedenen Bevölkerungsgruppen, aber auch Lebensmittel-Richtlinien bzw. -Verordnungen, wie Haltbarkeit, Handelsklassen, Kennzeichnungs-Verordnungen, erlaubte bzw. verbotene Zusätze, Qualitäts-Richtlinien, usw. Vorschläge dazu sind vorhanden, doch die Realisierung solcher notwendigen umfassenden Fakten-Datenbanken, wie sie bereits in anderen Forschungsgebieten vorhanden sind, ist auf dem Ernährungsgebiet noch nicht geschehen.