

Zur Erfassung und Verarbeitung medizinischer Daten mittels Computer

2. Mitteilung

Die Fehlerprüfung der durch das Datenerfassungs- und Speicherprogramm (DUSP) gespeicherten Daten

Aus der nuklearmedizinischen Abteilung des Evangelischen Krankenhauses Bethesda, Duisburg (Chefarzt: Dr. med. habil. H. A. E. Schmidt) und der Deutschen Klinik für Diagnostik, Wiesbaden

W. GIERE

Die Möglichkeiten der formalen Fehlerkontrolle, die sich aus der Art der Datenerhebung mit dem Datenerfassungs- und Speicherprogramm (DUSP) ergeben, werden beschrieben. Die Festlegbarkeit »zulässiger« Inhalte durch den Text der Erhebungsbögen erlaubt wirksame Prüfungen auf Fehler, d. h. »unzulässige« Inhalte. Die Standardisierbarkeit der Kontrollroutinen gestattet weitgehende Programmvereinfachungen. Wegen der Benutzung von Standardunterprogrammen kann sich die Programmierung auf die Definition der »zulässigen« Inhalte beschränken.

COLLECTION AND PROCESSING OF CLINICAL DATA BY COMPUTER

2. The Checking Program for Data Stored by the Data Collection and Storage Program (DUSP)

The possibilities of a formal data check corresponding to the Data Collecting and Storage Program (DUSP) are described. The definition of »allowed« contents by the text of the questionnaires permits effective search for input errors, i.e. not »allowed« contents. The standardizing of control routines widely allows to simplify the programs. By use of standard subroutines programming can be limited to the definition of the »allowed« contents.

1. Einleitung

Die erste Mitteilung über das Datenerfassungs- und Speicherprogramm (DUSP) (1) behandelte den Weg der Patientendaten von der Erhebungsstelle, z. B. dem Schreibtisch des Arztes, bis zur Speicherung durch den Computer auf der Magnetplatte.

Es sei erlaubt, einiges ins Gedächtnis zu rufen, was zum Verständnis der zu beschreibenden Fehlerkontrolle nötig scheint:

1.1 Die zu prüfenden Patientendaten stehen als variabel lange Datenblöcke auf der Magnetplatte. Diese Blöcke

entsprechen logisch zusammengehörenden Datengruppen, z. B. Adresse, Anamnese, Befunde usw. Jeder dieser Blöcke heißt FORMULAR.

1.2 Jedes FORMULAR ist vom nächsten FORMULAR durch ein Sonderzeichen, das Formular-Ende-Zeichen (FEZ), getrennt und durch Patientenummer (PNR) und Formular-Auswahl-Zeichen (FAZ) eindeutig identifiziert.

1.3 Jedes FORMULAR ist durch Sonderzeichen in variabel lange Sätze unterteilt. Ein einzelner Satz wird ZEILE genannt. Die Zahl der ZEILEN pro FORMULAR ist durch das Programm nicht festgelegt.

- 1.4 Zu Kontrollzwecken verlangt das DUSP, daß jede sechste ZEILE leer bleibt, d. h. daß zwei Sonderzeichen zur Zeilentrennung sich unmittelbar folgen.
- 1.5 Jede ZEILE (mit Ausnahme jeder sechsten — siehe 1.4) *muß* einen Inhalt (INH) und *kann* einen Zusatz (ZUS) enthalten. Durch ein Trennzeichen ist der INH vom ZUS eindeutig trennbar. Wenn ein ZUS vorhanden ist, folgt er *nach* dem INH. Es sei daran erinnert, daß die Einrichtung des ZUS der Möglichkeit klartextlicher Erläuterungen zum abgespeicherten INH dient. Auf das Wesen des INH wird im Zusammenhang mit der Fehlerkontrolle weiter unten noch einmal genauer einzugehen sein.
- 1.6 Vor der Speicherung prüft das DUSP die Gültigkeit der Patientenummer, die Zulässigkeit des gewählten Formular-Auswahl-Zeichens (FAZ), die korrekte ZEILEN-Zählung durch Kontrolle der in 1.4 genannten Bedingung, daß jede sechste ZEILE leer sein muß.
- 1.7 Die Daten werden per Lochstreifen in die Datenverarbeitungsanlage eingelesen. Der Lochstreifen entsteht beim Abschreiben spezieller Erhebungsbögen auf einer Schreibmaschine mit angeschlossenem Lochstreifenstanzer. Dabei entspricht einer Schreibmaschinenzeile eine ZEILE, einem Wagenrücklauf der Schreibmaschine die Kodierung für das Sonderzeichen der ZEILEN-Trennung, den von einem einheitlichen Erhebungsbogen abgeschriebenen Daten das FORMULAR und das Sonderzeichen für »Erhebungsbogenwechsel« dem Formular-Ende-Zeichen FEZ.
- 1.8 Ein Erhebungsbogen selbst besteht aus einem Textteil, der vorgedruckt ist, und einem frei gelassenen Kodierrand. Nur der letztere wird ausgefüllt und von der Sekretärin abgeschrieben. Zeilennummern werden nicht geschrieben.
- 1.9 Durch den Textteil des Erhebungsbogens läßt sich für jede Schreibzeile des Kodierrandes angeben, was die dort stehenden Zeichen bedeuten; es läßt sich aber auch eindeutig festlegen, welche Zeichen dort erscheinen dürfen. Das heißt: Der zugehörige Text in einer Zeile des Erhebungsbogens bestimmt, welcher INH in der korrespondierenden ZEILE des entsprechenden FORMULARS »zulässig« ist.

2. DUSP-spezifische Fehlermöglichkeiten

Der Untersucher füllt den Kodierrand des Erhebungsbogens mit den für den jeweiligen Patienten zutreffenden Angaben aus. Meist wird ihm die Befundung durch vorgeschlagene Abkürzungen erleichtert. Die Sekretärin schreibt die Eintragungen ab, dabei entsteht ein Lochstreifen, der in den Computer eingelesen wird. Dabei wird zwar die Formularidentifikation (vgl. 1.6), nicht aber die einzelne Eintragung geprüft. So kann es zu fehlerhaften INH kommen. Im einzelnen kann ein Fehler folgende Ursachen haben:

- 2.1 Falsches Kodieren durch den Untersucher. Diese Fehler können durch eine formale Prüfung nur erkannt werden, wenn der falsche Kode »unzulässig« ist. Da die Kodierweise selbst aber durch den Text des Erhebungsbogens genau vorgeschrieben ist, dürften diese Fehler sehr selten sein. Sie sind sicher seltener als bei numerischen Verschlüsselungsmethoden, wohl auch seltener als bei der Verwendung von Markierungsbelegen, bei denen man sich doch u. U. einmal in der Markierungsstelle irrt. Verschlüsselungs- und Markierungsfehler lassen sich im Gegensatz zu den geschilderten Kodierfehlern nur schwer prüfen.
- 2.2 Falsches Lesen durch die Sekretärin. Diese Fehler werden in der Regel zu einem »unzulässigen« Kode führen und damit formal prüfbar.

2.3 »Vertippen« der Sekretärin. Bei Blindschreiben werden relativ häufig neben- oder übereinander auf der Schreibmaschinentastatur angeordnete Zeichen verwechselt. Sofern diese Verwechslungen zu einem »unzulässigen« Kode führen, sind auch diese Fehler durch formale Prüfungen feststellbar.

2.4 Der INH der einzelnen ZEILEN stimmt deswegen nicht, weil das FORMULAR ein falsches Formular-Auswahl-Zeichen (FAZ) trägt: das heißt, daß es sich zwar um ein an sich gültiges FAZ handelt (sonst wären die Daten gar nicht erst gespeichert worden), daß es aber mit dem eines anderen Erhebungsbogens verwechselt wurde, so daß die Spezifikationen für die einzelnen ZEILEN nicht zu den gespeicherten INH passen.

Dieser Fehler läßt sich durch formale Prüfungen gut erkennen, wenn man die verschiedenen Erhebungsbögen in ihrer Gestaltung möglichst weitgehend variiert — was sich meist ohnehin aus der Sache ergibt.

3. Aufgabenstellung

Im Folgenden soll nur noch von *der* Fehlerprüfung die Rede sein, die rein formal prüft, ob der INH im Hinblick auf den korrespondierenden Text des Erhebungsbogens »zulässig« ist. Es soll mit anderen Worten weder auf Plausibilitätsprüfungen (beispielsweise Männer mit Menopause, Frauen mit Prostatahypertrophie, Meßwerte in unmöglichen Größenordnungen usw.) eingegangen werden, noch auf die Prüfbarkeit des ZUS, der ja als Ergänzung des INH in jeder ZEILE erlaubt ist. Da diese Möglichkeit der klartextlichen Ergänzung des INH durch einen ZUS gerade für die seltenen und atypischen Befunde geschaffen wurde, welche durch den Text der Erhebungsbögen *nicht* vordefiniert sind, entzieht sich der ZUS einer Prüfung im oben geschilderten Sinne.

Die formale Prüfung der DUSP-gespeicherten Daten erfordert also erstens zeilenweises Festlegen der »zulässigen« INH, zweitens zeilenweises Prüfen, ob keine anderen als die »zulässigen« INH durch DUSP gespeichert wurden. Diese Fehlerprüfung soll durch den Computer im Anschluß an die Speicherung durch ein Prüfprogramm erfolgen. Als fehlerhaft erkannte Daten werden dabei ausgedruckt und können verbessert werden (vgl. 1. Mitteilung).

4. Standardisierung der »zulässigen« INH

Für die Programmierung dieser Art von Fehlerprüfung erweist es sich als Vorteil, daß die vorkommenden INH sich formal in Gruppen einteilen und standardisieren lassen. Die so entstandenen Gruppen nennen wir INH-TYPEN. Wir unterscheiden »unkodierte« und »kodierte« INH-TYPEN: »Unkodierte« sind solche, deren INH aus unverschlüsselten Antworten auf die Fragen der Erhebungsbögen bestehen; als »kodierte« haben wir alle die INH bezeichnet, die verschlüsselte Daten enthalten. Dabei kann die Art der Verschlüsselung unterschiedlich sein. Die wichtigsten dieser INH-TYPEN seien kurz vorgestellt:

- 4.1 INH **unkodiert alphanumerisch** — TYP IUA. INH ist alphanumerischer Klartext, z. B. Medikamentennamen, Adressen, Hauptbeschwerden usw. TYP IUA ist auf dem Erhebungsbogen überall dort vorgesehen, wo sich eine Verdichtung des Informationsgehaltes nur schlecht oder unter den gegebenen Umständen gar nicht erreichen läßt. Die Länge des INH ist beim TYP IUA nicht definiert.
- 4.2 INH **unkodiert numerisch** — TYP IUN. INH ist eine unverschlüsselte numerische Angabe, ein Meßwert z. B., das Ergebnis einer Zählung als Antwort auf die

Frage »Wie oft?« im Erhebungsbogen usw. Weder die Größe der als INH gespeicherten Zahl, noch die Länge des INH sind standardmäßig festgelegt.

4.3 INH kodierte Zeitangabe — TYP IKZ. INH ist eine nach den Regeln des DUSP kodierte Angabe über die Zeitdauer: Sie besteht aus ein oder zwei Ziffern, unmittelbar gefolgt von einem Buchstaben. Die Größenordnung der Zahl vor dem Buchstaben ist festgelegt auf 1 bis 99; die »zulässigen« Buchstaben beschränken sich auf die Abkürzungen s, t, w, m, j für die Zeiteinheiten Stunde, Tag, Woche, Monat, Jahr. »Drei Monate« als Zeitangabe sähe also im TYP IKZ »3m« aus, »Siebzehn Tage« entspräche »17t«. Sowohl die Zahl der zulässigen Zeichen, als auch die Länge des INH sind begrenzt.

4.4 INH kodiert medizinisch — TYP IKM. »Medizinisch« wurde die recht ungenaue subjektive Skalierung in eine negative und drei positiven Klassen genannt: nein = 0, leicht = 1, mittel = 2, schwer = 3 oder: o. B. = 0, fraglich = 1, deutlich = 2, stark = 3 usw. Bei TYP IKM sind vier Ziffern zulässig, die Länge des INH ist auf ein Zeichen beschränkt.

4.5 INH Buchstabenkode. Hierbei enthält der INH Abkürzungen für Worte, die auf dem Erhebungsbogen in Form eines sog. »multiple choice«, einer Auswahlfrage, vorgesehen sind, beispielsweise »w« für »weich«, »p« für »pulsierend«, »r« für »rechts«, »e« für »euthyrot« oder »e« für »Eklampsie«. Aus dem Text des Erhebungsbogens ergibt sich der Sinn der Buchstaben; außerdem sind die als Abkürzung »zulässigen« gekennzeichnet.

Bei dieser Methode der Kodierung müssen im Hinblick auf die Fehlerprüfung zwei verschiedene Typen unterschieden werden. Sie entsprechen erstens der Auswahlfrage mit nur einer erlaubten Antwort, zweitens der mit mehreren erlaubten Antworten.

4.5.1 INH kodiert 1 Antwort — TYP IK1. INH kann nur ein Buchstabe sein, die Zahl der »zulässigen« Buchstaben hängt von der Formulierung des Erhebungsbogens ab.

Als Sonderfall ist es bei diesem TYP IK1 erlaubt, in einer ZEILE mehrere Auswahlfragen zu stellen, für die je eine Antwort erlaubt ist. Bis zu drei »multiple choice«-Fragen pro Zeile sind vorgesehen, dabei erhöht sich die Zahl der Zeichen im INH entsprechend der Zahl der Fragestellungen auf bis zu drei Zeichen. Bei dieser Art der Kodierung lassen sich also die einzelnen Zeichen des INH immer exakt der ersten oder zweiten oder dritten Frage zuordnen: Die einzige mögliche Antwort auf die zweite Frage ist durch den zweiten Buchstaben im INH ausgedrückt.

Ein Beispiel soll dies verdeutlichen: Zur Beschreibung der Lokalisation könnte nach drei Dimensionen in einer Zeile des Erhebungsbogens gefragt sein.

»Kranial/Caudal, Dorsal/Ventral, Rechts/Mitte/Links«. Die drei Fragestellungen sind durch Kommata getrennt; die Schrägstriche kennzeichnen die alternativen Antwortmöglichkeiten; die Unterstreichungen zeigen die »zulässigen« Buchstaben. Hieraus ergibt sich, daß an erster Stelle k oder c, an zweiter Stelle d oder v, an dritter Stelle r, m oder l stehen darf. Wird eine der Fragen nicht beantwortet, muß — entsprechend den Regeln des DUSP (vgl. 1. Mitteilung) — das Zeichen für »keine Angabe« eingefügt werden.

Die drei möglichen Stellen können getrennt geprüft werden. Es handelt sich dabei um eine mehrfache Anwendung des TYP IK1. Es ist Zufall, nicht Notwendigkeit, daß im hier angeführten Beispiel unterschied-

liche Buchstaben gewählt wurden. Erforderlich ist nur, daß sich die Buchstabenkodes innerhalb einer einzelnen Fragestellung unterscheiden.

Zusammenfassend läßt sich zum TYP IK1 formulieren: Ein Buchstabe pro Fragestellung, maximal drei Buchstaben pro INH, entsprechend höchstens drei Fragestellungen pro Zeile des Erhebungsbogens; jeder Buchstabenkode wird unabhängig von den anderen geprüft.

4.5.2 INH kodiert viele Antworten — TYP IKV. Bei mehreren erlaubten Antworten pro Auswahlfrage ist die Zahl der möglichen Zeichen im INH maximal so groß wie die Zahl der Antworten, die auf dem Erhebungsbogen als Auswahl formuliert wurden. Das heißt, die Länge ist beim TYP IKV — anders als beim TYP IK1 — variabel. Außerdem ist die Reihenfolge der Buchstabenkodes beliebig, da der ersten, zweiten, dritten usw. Stelle im INH keine Bedeutung für die formale Fehlerkontrolle zukommt. Es wird lediglich geprüft, ob sich im vorliegenden INH »unzulässige« Kodes befinden. Sonst gelten dieselben Spezifikationen wie für den TYP IK1.

Die Bedeutung der Buchstaben ergibt sich allein aus dem zugeordneten Text der Auswahlfrage.

5. INH-TYP-spezifische Prüfungen

Rein formal lassen sich fast alle der genannten INH-TYPEN mit dem Ziel prüfen, »unzulässige« INH festzustellen. Lediglich der TYP IUA bietet hierzu keine Möglichkeit, da er jede alphanumerische Angabe ohne Einschränkung zuläßt.

Die Prüfungsmöglichkeiten für die anderen geschilderten TYPEN sind im einzelnen folgende:

5.1 IUN: Prüfung auf Zahlen, Länge nicht prüfbar.

5.2 IKZ: Prüfung auf Zahl von 1 bis 99 gefolgt von einem der Buchstaben s, t, w, m oder j; Prüfung auf Länge: minimal 2, maximal 3 Zeichen.

5.3 IKM: Prüfung auf Ziffer 0, 1, 2, 3; Prüfung auf Länge: 1 Zeichen.

5.4 IK1: Prüfung auf »zulässige« Buchstaben; Prüfung auf Länge: 1 Zeichen pro Fragestellung, d. h. minimal 1, maximal 3 Zeichen pro INH.

5.5 IKV: Prüfung auf »zulässige« Buchstaben; Länge nicht prüfbar.

6. Programmierung

Für die genannten Prüfungen wurden Standardunterprogrammen entwickelt, die für jeden INH-TYP alle Prüfmöglichkeiten ausschöpfen. Zur Anwendung muß erstens die dem jeweiligen INH-TYP entsprechende Standardroutine aktiviert, zweitens mit Angaben über die »zulässigen« Zeichen spezifiziert werden. Hierfür wurden Befehle entwickelt, die auch Laien nach kurzer Einarbeitungszeit das »Programmieren« eigener Prüfprogramme gestatten. Da wir die Prüfprogramme aus Gründen der Maschineneitersparnis immer mit weiterverarbeitenden Programmen koppelten, zum Beispiel mit der Ausgabe programmierter Arztbriefe, sollen diese Befehle bei der Beschreibung des diesen Programmen zugrunde liegenden Dekodierungs- und Textausgabe-Programmsystem (DUTAP) mitbesprochen werden. Die Beschreibung des DUTAP und seiner Anwendung muß einer weiteren Mitteilung vorbehalten bleiben.

7. Wirksamkeit der beschriebenen Prüfungen

Die Wirksamkeit der Prüfungen ist schwer abzuschätzen. Generell ist sie umgekehrt proportional zur Vielfalt der »zulässigen« INH, also bei IK1 wirksamer als bei IKV, bei

IKZ wirksamer als bei IUN. Da es bei IKM nur vier richtige Antworten gibt, sollte man meinen, TYP IKM besonders wirksam prüfen zu können. Bedenkt man aber, daß 1, 2 und 3 auf der Schreibmaschine nebeneinander liegen, dürfte die Prüfung wegen der Möglichkeit des »Vertippens«, ohne daß ein »unzulässiger« Kode entsteht, doch nicht ganz zufriedenstellend sein.

Immerhin läßt sich soviel sagen, daß wir in Stichproben von wöchentlich mindestens 20 FORMULAREN über ein halbes Jahr in der nuklearmedizinischen Abteilung des Evangelischen Krankenhauses Bethesda in Duis-

burg keinen Fehler gefunden haben, der die geschilderten formalen Kontrollen unentdeckt passiert hätte.

Literaturverzeichnis

- (1) GIERE, W. und BAUMANN, H.: Zur Erfassung und Verarbeitung medizinischer Daten mittels Computer. 1. Mitteilung: Ein Datenerfassungs- und Speicherprogramm (DUSP) zur Dokumentation von Krankengeschichten. Meth. Inform. Med. 8: 11—18, 1969.

Anschrift des Verfassers: Dr. med. Wolfgang Giere, Deutsche Klinik für Diagnostik A.G., z. Zt. 6000 Frankfurt, Dantestraße 5.