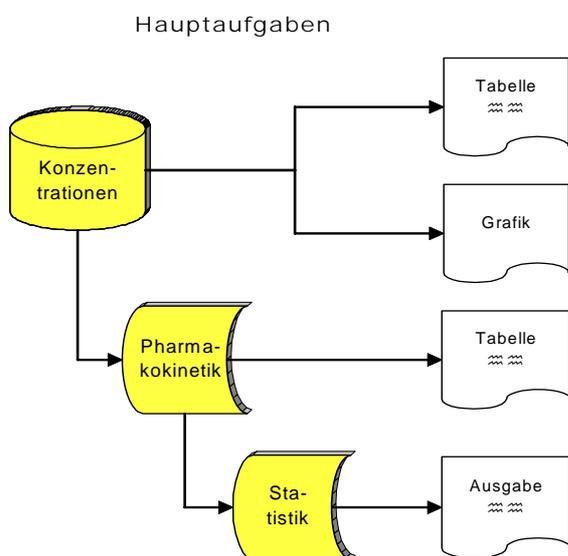


Ein System von SAS-Programmen zur Auswertung von Pharmakokinetik-Studien

Günther Pabst, L.A.B. GmbH & Co, Neu-Ulm

Zusammenfassung: Ein System von SAS-Programmen und Makros wurde entwickelt zur standardisierten, aber dennoch flexiblen Auswertung von Studien zur Pharmakokinetik und relativen Bioverfügbarkeit bzw. Äquivalenz von Arzneimitteln. Die Ausgabe erfolgt generell in Form von ASCII-Dateien, die ohne manuelle Nachbereitung unmittelbar in den Bericht übernommen werden können.

Die Pharmakokinetik beschreibt, was mit einem Arzneimittel im Körper geschieht, von der Aufnahme (Absorption) über die Verteilung (Distribution) bis zum Abbau (Elimination) durch Metabolismus oder Ausscheidung. Um die Pharmakokinetik eines Arzneimittels zu beschreiben, werden Blutproben nach einer einmaligen oder mehrmaligen Gabe des Präparates abgenommen und auf ihre Konzentrationen der fraglichen Substanz hin analysiert.



Die Auswertung derartiger Daten eignet sich hervorragend für eine Standardisierung, denn die Datenstruktur bleibt gleich, auch wenn sich Substanz- und Präparatebezeichnungen ändern. SAS wurde als Auswertesystem gewählt, da vorab entschieden war, daß die statistische Auswertung unter SAS vorgenommen werden sollte. Bei der Entwicklung des Auswertesystems sollten vorrangig die folgenden Ziele erreicht werden:

- Jede Benutzereingabe von Daten oder Texten sollte möglichst nur einmal erforderlich sein – beispielsweise soll die Beschreibung der Substanz oder Behandlung aus dem Hintergrund eingefügt werden, gleichgültig ob es sich um Tabellen, Grafiken usw. handelt.
- Die Ausgabe sollte in einer Form erfolgen, die ohne weitere Nachbearbeitung unmittelbar für einen Bericht verwendet werden kann – z.B. soll für vordefinierte Anmerkungen an Daten automatisch eine erklärende Fußzeile erzeugt werden, falls eine derartige Annotation in einer Tabelle vorkommt, aber auch nur dann. Anmerkung: Aufgrund seines Einsatzspektrums erfolgen alle Ausgaben in englischer Sprache.
- Die Ausgabe soll gleichzeitig, falls erforderlich, eine manuelle Nachbereitung erlauben.
- Das System soll bei einer Fehlbedienung mindestens eine Fehlermeldung abgeben, besser eine Fehlbedienung gar nicht erst zulassen – z.B. soll eine tabellari-

sche Anlistung von Daten einer Behandlung nur angefordert werden können, wenn derartige Daten überhaupt vorliegen.

- Das System soll benutzerfreundlich sein, wobei die Funktionalität wichtiger ist als die grafische Oberfläche. Wenn der Benutzer z.B. an einer Stelle unter mehreren Optionen auswählen kann, sollen diese alle am Bildschirm gleichzeitig angezeigt werden. Die Auswahl kann dann durch Eingabe erfolgen; es wurde nicht als erforderlich erachtet, eine Auswahl durch 'Anklicken' zu ermöglichen.
- Das System soll trotz aller Standardisierung möglichst variabel sein, indem es viele Optionen anbietet, aber auch leicht weitere Modifikationen erlauben, falls solche erforderlich sein sollten.

Voraussetzung für ein derartig standardisiertes System ist natürlich, daß die Daten in einer fest vorgegebenen Struktur (in SAS DATA SETs) vorliegen. Beispielsweise gehört zu jedem Meßwert eine 1-Charakter Variable, die ggf. Codes für Sonderfälle enthält, z.B. "!" = Wert eingeschränkter Zuverlässigkeit: in Tabellen so zu annotieren, aber in Auswertungen usw. wie ein voll gültiger Wert. Bei unquantifizierbar niedrigen Konzentrationen wird als „Konzentration“ die zutreffende untere analytische Bestimmungsgrenze eingetragen, die sich durchaus von einer Probe zur anderen unterscheiden kann, im Codefeld steht dann "<", für Berechnungen ist dies wie eine Null zu berücksichtigen.

Die Benutzerführung und (hoffentlich) -freundlichkeit soll hier mit den Schritten erläutert werden, die erforderlich sind, um eine Tabelle der Konzentrationsdaten zu erzeugen. Standardeinstellung ist dabei: Je eine Tabelle pro Behandlung, Spalten = Probennahmezeitpunkte, Zeilen = Werte eines Teilnehmers.

1. Angabe der Substanz (Kürzel). Diese Kürzel wurden vorher definiert, wobei u.a. ein beschreibender Text anzugeben war, der bei jeder Ausgabe ausgedruckt werden soll, z.B. „Konzentrationen von Nifedipin im Plasma“. Am Bildschirm werden diejenigen vordefinierten Kürzel gezeigt, für die auch tatsächlich Daten vorliegen – und auch nur unter diesen kann ausgewählt werden.
2. Angabe der Behandlungen (Kürzel), für die eine Tabelle erstellt werden soll. Die Kürzel wurden gleichfalls in einem vorausgegangenen Arbeitsschritt definiert – mit bis zu zwei Textzeilen zur Beschreibung der jeweiligen Behandlung. Im Menue werden nur die Kürzel derjenigen Behandlungen zur Auswahl angeboten, für die tatsächlich Daten vorliegen. Dieser Schritt dient gleichzeitig zur Eingabe der Tabellennummern (Daten ohne Tabellenummer bzw. Tabellenummer ≤ 0 werden nicht ausgegeben). Auf diese Weise ist eine von Behandlung zu Behandlung unterschiedliche Formatierung der Ausgabe möglich.
3. Angabe der gewünschten Spaltenbreite, 8 bis maximal 12 Druckpositionen einschl. der Abstände zwischen den Spalten.
4. Angabe der gewünschten deskriptiven Statistik über die Daten zu einem Probenahmezeitpunkt. Zur Auswahl stehen: Fallzahl (n), Minimum, Median, Maximum, Mittelwert, Standardabweichung, Standardfehler, Variationskoeffizient, geometrisches Mittel, geom. Streufaktor und geom. Variationskoeffizient – es ist aber auch möglich, eine Tabelle ohne derartige Statistik zu erzeugen.
5. Auswahl des Layouts, z.B. mit/ohne Angabe der individuellen Behandlungsreihenfolge. Das System errechnet, ob und wie viele Seitenumbrüche erforderlich sind bei Druck im Hochformat (Portrait-Mode) oder Querformat (Landscape-Mode).
6. Auswahl des Druckformats der Einzeldaten. Angeboten wird bestmöglich (z.B. BEST6.), Fixformat, z.B. 6.2 (wobei nur so viele Dezimalstellen angefordert

werden können, daß auch der größte Einzelwert noch ohne Rückfall auf das BEST-Format gedruckt werden kann) und schließlich Druck mit einer festen Anzahl signifikanter Stellen, denn Analytikergesulte sind in der Regel nur auf z.B. $\pm 5\%$ genau und es ist nicht sinnvoll, mehr als drei signifikante Ziffern anzugeben. Das System schlägt aufgrund der Datenlage ein geeignetes Fixformat vor. Die Formatauswahl wird permanent gespeichert, um einerseits ggf. erforderliche Neuausdrucke der gleichen Tabelle zu erleichtern, andererseits weil das Format auch Anwendung findet auf pharmakokinetische Kenngrößen wie z.B. C_{max} , die maximale beobachtete Konzentration.

↳ den Rest macht das Auswertesystem!

Beispiel (Daten aus dem Hintergrund, Benutzereingabe, vom Auswertesystem beigefügt falls erforderlich)

Table 1 (cont.)

Data: Concentrations of Nifedipine in plasma [ng/ml]

A: Test preparation, 10 mg p.o.

Subject	predose	15min	30min	45min	1h	1.5h
1	<0.5	<0.5	15.0	22.1	32.2	31.3
2 N	<0.7	5.13	9.61	18.3	22.1	20.5!
3	<0.5	7.23	14.5	33.7	32.8	25.3
.
.
.
12	<0.5	11.1	22.9	25.3	21.2	17.5
n	11	11	11	10	11	11
mean	0.00	8.23	15.44	22.56	28.31	25.74
%CV	0.0%	19.1%	14.5%	17.3%	21.1%	34.6%
geomean	.	7.84	14.46	20.21	25.44	21.22

N : data of this subject not used for descriptive statistics
! : value of limited reliability

Hierzu die folgenden Anmerkungen:

- Spaltenüberschriften wurden bereits bei der „Definition“ der Substanz festgelegt.
- Werte kleiner als die Bestimmungsgrenze werden als z.B. "<0.5" gelistet, es gibt aber auch die Option, hierfür eine feststehende Abkürzung zu verwenden, z.B. "BLQ" mit entspr. Erläuterung in einer Fußzeile.
- Fußzeilen für vordefinierte Annotationen, die in der Tabelle nicht vorkommen, z.B. "*" : value not reasonable, not used", werden unterdrückt.
- Bei horizontalem Seitenumbruch, z.B. wegen zu großer Zeilenzahl bei Verwendung von Querformat, werden die statistischen Kenngrößen nur auf der letzten Seite gedruckt.
- Die statistischen Kennzahlen werden intern über PROC MEANS errechnet, der Median, falls gefordert, über PROC UNIVARIATE, das geometrische Mittel über Log-Transformation - PROC MEANS - Rücktransformation. Diese Kennzahlen werden mit unterschiedlicher Formatierung ausgedruckt: Minimum, Maximum formatiert wie die Daten, n als ganze Zahl, Variationskoeffizienten mit 1 Dezimalstelle, alle anderen, falls möglich, mit einer Stelle mehr als wie bei den Rohdaten. Bei der Berechnung geometrischer Kennzahlen bleiben Werte ≤ 0 unberücksichtigt.

Deshalb wird automatisch eine weitere Zeile ausgegeben mit der ggf. reduzierten Fallzahl für diese Kenngrößen.

- Die Ausgabe erfolgt in eine ASCII-Datei (mit Leerzeichen und 'Form Feed'), die damit unmittelbar auf den Drucker geschickt werden kann, jedoch auch einer manuellen Nachbereitung zugänglich ist. Für Grafiken werden ASCII-Dateien geschrieben, die HPGL-Code enthalten, was einerseits eine Nachbereitung erlaubt, andererseits auch erlaubt, derartige Grafiken in die Textverarbeitung zu integrieren.

Die trotz der vielfachen Optionen ggf. noch erforderliche Variabilität wurde erreicht, indem das Auswertesystem in Module mit SAS-Programmen gegliedert wurde, die mit %INCLUDE geladen und ausgeführt werden. Die für den Benutzer sichtbare „Oberfläche“ im Program Editor für das o.g. Beispiel sieht folgendermaßen aus:

PROGRAM EDITOR - ctp rint.sas		-	□	×
Command ==>				▲
00001 * ctp rint.sas ----- Version 2.01 21JUN96				
Pb;				
00002 *				
00003 * ctp rint.sas = concentration tables print				
00004 *				
-----;				
00005 * Generates concentration tables (one per treatment)				
00006 * as ASCII-files 'ct*' where * is the table number				
00007 *				
-----;				
00008 *;				
00009 * for Plasma, serum, blood use: PL, for Urine data use: U;				
00010 %LET matrix=PL;				
00011 *;				
00012 * ----- to be repeated for each substance/substrate -----;				
00013 OPTIONS NOSOURCE;				
00014 DM 'LOG; CLEAR;';				
00015 FILENAME pgm LIST;				
00016 * ----- Substance/Substrate to be printed;				
00017 %select (DAT,1);				
00018 %INCLUDE pgm(ctp rint2.sas);				
00019 * ----- descriptive statistics;				
00020 %INCLUDE pgm(ctp rint3.sas);				
00021 %INCLUDE 'sas.sas';				
00022 * ----- print format;				
00023 %INCLUDE pgm(ctp rint4.sas);				
00024 * ----- create macro invocation;				
00025 %INCLUDE pgm(ctp rint5.sas);				
00026 %INCLUDE 'sas.sas';				
00027 * autocall macros ctp r and formd needed;				
00028 *;				
00029 PROC DATASETS MT=DATA LIB=WORK NOLIST;				
00030 DELETE s test;				
00031 RUN;				
00032 QUIT;				
00033 X 'del sas*.sas';				
00034 OPTIONS SOURCE;				
				▼
◀			▶	

Von Verfügbarkeit der Daten und/oder Auswerteoption abhängige Verzweigungen wurden z.T. dadurch realisiert, daß SAS Code in Dateien 'sas.sas' geschrieben wird, die dann gleichfalls mit %INCLUDE zur Ausführung kommen. Die Datei 'sas.sas', die in Zeile 21 aufgerufen wird, wurde z.B. von ctp rint3.sas erzeugt und enthält SAS Code zur Berechnung der deskriptiv statistischen Kennzahlen. Falls erforderlich, kann so der Code leicht angepaßt werden. Es natürlich klar, daß bei einem derartigen Eingriff die Validierung des Systems verloren geht.

Bei der Programmierung des Auswertesystems in SAS wurde auf die Verwendung von SAS/AF verzichtet, um einfacher eventuell dennoch erforderliche Eingriffe vornehmen zu können. Desgleichen wurden MACROs nur dann verwendet, wenn gleichartige Teile in mehreren Ablaufschemata vorkamen. Bei der Grafik wurde auf SAS/Graph verzichtet, zum einen wegen des Zeitbedarfs bei der Ausführung (zumindest als mit der Entwicklung des Auswertesystems begonnen wurde; Seriengrafiken von 24 Probanden mit je drei Kurven in einem Bild anzufertigen erfordert im derzeitigen Stand ca. 1 Minute Arbeitszeit), zum anderen wegen des Platzbedarfs (in unserem System wird ein ASCII-File mit HPGL-Code erstellt, der im o.g. Beispiel ca. 3 kB groß ist), zuletzt aber auch wegen der hohen Lizenzgebühr.

Das vorgestellte Auswertesystem wurde im Verlauf mehrerer Jahre entwickelt, zuerst in SAS für DOS; hier wurde jedoch die erweiterte Version für Windows vorgestellt. Das System enthält mittlerweile Programme zu praktisch allen Auswerteschritten, z.B. auch zur Baseline-Korrektur oder für Urindaten. Für die statistische Auswertung stehen auch nonparametrische Verfahren zur Verfügung. Insgesamt besteht das System deshalb z.Zt. aus über 100 Modulen und Macros.