

Datenauswertung mit SAS unter Verwendung der LAG-Funktion

Bernd Jäger^{*}, Paul Eberhard Rudolph^{**}, Karl-Ernst Biebler^{*},
Andreas Greinacher^{***}

^{*} Institut für Biometrie und Medizinische Informatik,
Ernst-Moritz-Arndt-Universität Greifswald

^{**} Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere Dummerstorf/Rostock

^{***} Institut für Immunologie und Transfusionsmedizin,
Ernst-Moritz-Arndt-Universität Greifswald

Kurzfassung

Die Auswertung klinischer Daten selbst mit einfachen statistischen Prozeduren kann sich mitunter als relativ kompliziert erweisen, wenn die entsprechenden Daten nicht in einer für das Auswertungsverfahren adäquaten Form vorliegen. Das kann z. B. bedeuten, daß vorhandene Beobachtungen für einen Patienten erst ergänzt werden müssen unter Zurückgreifen auf bereits vorliegende Beobachtungen am gleichen Patienten.

Dazu stellt SAS mit der LAG-Funktion ein geeignetes Werkzeug zur Verfügung. Die Verwendung der LAG-Funktion wird an dem folgenden Beispiel aus der Medizin demonstriert.

In zwei Studien wurden Patienten, die heparininduzierte Antikörper gebildet hatten, auf ein alternatives Medikament (Hirudin) umgestellt, das als Dauerinfusion über einen Zeitraum von ca. zehn Tagen verabreicht wurde. Die Infusion, deren einziger Stellparameter die Infusionsgeschwindigkeit ist, wurde nur unterbrochen, wenn es zu Behandlungskomplikationen kam. Es sollte in der Studie unter anderem die Frage geklärt werden, ob die empfohlene Anfangsdosis der Hirudininfusion zu hoch oder zu niedrig angesetzt wurde und während der Therapie im wesentlichen nach unten oder oben korrigiert werden mußte. Von über 200 Patienten liegen für diese Fragestellung bis zu 52 Datensätze pro Patient vor. Jeder Datensatz enthält Angaben zum Infusionsstart, zum Infusionsende und zur Dosierung (Infusionsgeschwindigkeit).

Zur Beantwortung der o. g. Frage soll die Dosierungsfunktion, eine Treppenfunktion, durch eine Regressionsgerade nachgebildet werden. Als problematisch erweist sich dabei, daß Behandlungslücken aus zwei aufeinanderfolgenden Infusionen erst in der Datei erkannt werden müssen, weil sie in der Studie nicht als solche protokolliert wurden. Unter Verwendung der LAG-Funktion wurden die vorhandenen Datensätze deshalb durch die nichtprotokollierten Beobachtungen mit der Infusionsgeschwindigkeit Null ergänzt. Wird außerdem noch aus den vorhandenen Zeitangaben für jeden Datensatz eines Patienten die zur entsprechenden Infusionsgeschwindigkeit (INFSPEED) gehörende absolut verstrichene Zeit seit Start der Behandlung als weitere Variable (DIFFT) berechnet, können die Parameter der gewünschten Regressionsgeraden mit der SAS-Prozedur REG bestimmt werden.

Der Übersichtlichkeit halber werden Datensätze von nur drei Patienten präsentiert.

1. Einleitung

In zwei multizentrischen Studien wurden Patienten, die heparininduzierte Antikörper gebildet hatten, auf ein alternatives Medikament umgestellt. Es handelt sich um Hirudin, ein dem Antigerinnungswirkstoff des Blutegels ähnliches, auf gentechnologischem Wege erzeugtes Präparat mit einer kurzen Halbwertszeit im Organismus (1-2 Stunden). Deshalb wird es als Dauerinfusion über einen Zeitraum von ca. zehn Tagen verabreicht.

Die Hirudinwirkung wird anhand individueller, optimaler Zielkorridore von Blutgerinnungsparametern beurteilt, beispielsweise der aktivierten partiellen Thromboplastinzeit (APTT). Werden APTT-Werte oberhalb des Zielbereiches gemessen, wird die applizierte Dosis pro Zeit über die Infusionsgeschwindigkeit nach unten geregelt. Werden Werte unterhalb des

Zielbereichs gemessen, wird die Dosis gesteigert. So soll verhindert werden, dass es bei zu hohen APTT-Werten zu Blutungen bzw. bei zu niedrigen zu thromboembolischen Komplikationen kommt. Die Infusionsgeschwindigkeit ist der einzige Stellparameter, weil die Infusionslösung den Wirkstoff Hirudin in einer festen Konzentration enthält.

Die Infusion sollte nur unterbrochen werden, wenn es zu Behandlungskomplikationen kommt, beispielsweise Blutungen, oder wenn Operationen notwendig werden.

Es sollte in der Studie unter anderem die Frage geklärt werden, ob die empfohlene Anfangsdosis der Hirudininfusion zu hoch oder zu niedrig angesetzt wurde, unter der Therapie im wesentlichen nach unten oder oben korrigiert werden mußte. Zukünftige Eingangsdosisempfehlungen wären entsprechend abzuändern. Allerdings handelt es sich hierbei nicht um die klassische Dosisfindungsproblematik der Pharmakologie, die andere Methodiken erfordert.

2. Datenbasis und methodische Vorgehensweise

Von über 200 Patienten liegen für diese Fragestellung bis zu 52 Datensätze pro Patient vor. Jeder Datensatz enthält Angaben zum Infusionsstart (Datum=INFDATS, Uhrzeit=INFTIMS), zum Infusionsende (Datum=INFDATE, Uhrzeit=INFTIME) und zur Dosierung (Infusionsgeschwindigkeit=INFSPEED).

Ziel soll es sein, die Dosierungsfunktion, eine Treppenfunktion, durch eine Regressionsgerade nachzubilden. Das Vorzeichen des Anstiegs der Regressionsgeraden gibt Auskunft über die Dosisveränderungen. Ist das Vorzeichen positiv, deutet es auf einen Dosisanstieg hin, ist es negativ, spricht es für eine Reduktion in der gesamten Behandlungszeit (siehe Abb.1).

Problematisch erweist sich, dass Behandlungslücken aus zwei aufeinanderfolgenden Infusionen erst in der Datei erkannt werden müssen, weil sie in der Studie nicht als solche protokolliert wurden. Nachdem sie zusätzlich als Infusionen mit der Infusionsgeschwindigkeit Null in die Datei eingefügt wurden, ist die Prozedur REG aus SAS abarbeitbar, wenn als abhängige Variable die Infusionsgeschwindigkeit und als unabhängige Variable die absolut verstrichene Zeit seit Start der Behandlung gewählt wurden. Die absolute Zeit muß aus den Datums- und Zeitangaben berechnet werden.

In kleinen, nachvollziehbaren Schritten werden die einzelnen Programmschritte beschrieben. Ausgangspunkt ist folgende SAS-Datei mit 20 Datensätzen, die der besseren Übersichtlichkeit wegen die Angaben von nur drei Patienten enthalten. Die Patienten mit der Identifikationsvariablen ISUBJECT sind in Tabelle 1 ff. durch horizontale Linien voneinander getrennt. Bei den mit „*“ gekennzeichneten Datensätzen erkennt man eine Behandlungslücke. Der folgende, zum gleichen Patienten gehörende Datensatz schließt in einem solchen Falle nicht nahtlos an das vorhergehende Behandlungsintervall an. Ziel der nachfolgenden Programmschritte wird es sein, an diese Stellen neue „Behandlungsintervalle“ einzufügen, die mit entsprechenden Anfangs- und Endzeiten belegt und mit der Infusionsgeschwindigkeit Null versehen die Behandlungslücken schließen.

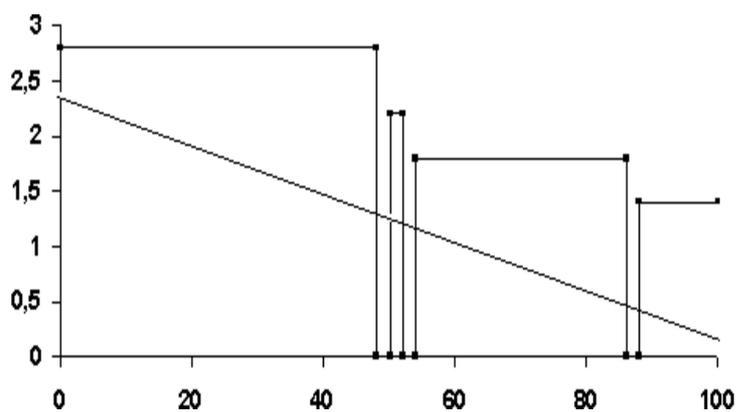


Abbildung 1: Treppenfunktion der Dosierungs-Zeit-Funktion und näherungsweise Beschreibung durch die zugehörige lineare Regressionsfunktion (Pat.Nr. 11601)

Tabelle 1: Ausgangsdatei ACH.HB1

ISUBJECT	INFDATS	INFTIMS	INFDATE	INFTIME	INFSPEED
10901	23APR94	2:05	23APR94	11:30	5.6
10901	23APR94	11:30	23APR94	20:30	5.6
10901	23APR94	20:30	24APR94	10:45	3.8
10901	24APR94	14:30	24APR94	23:00	5.6
10901	24APR94	23:00	25APR94	8:30	5.6
10901	25APR94	8:30	25APR94	18:00	5.6
10901	25APR94	18:00	26APR94	3:30	5.6
10901	26APR94	3:30	26APR94	13:00	5.6
10901	26APR94	13:00	27APR94	0:30	5.6
10901	27APR94	0:30	27APR94	10:00	5.6
10901	27APR94	10:00	27APR94	14:30	5.6 *
10901	27APR94	19:30	28APR94	5:00	5.6
10901	28APR94	5:00	28APR94	14:30	5.6
10901	28APR94	14:30	01MAY94	0:00	5.6
10901	01MAY94	0:00	04MAY94	8:00	5.6
<hr/>					
11201	19MAY94	12:00	28MAY94	12:10	4.9
<hr/>					
11601	03JUN94	22:00	05JUN94	22:00	2.8 *
11601	06JUN94	0:00	06JUN94	2:00	2.2 *
11601	06JUN94	4:00	07JUN94	12:00	1.8 *
11601	07JUN94	14:00	08JUN94	20:00	1.4

Die beschriebenen Aufgaben, Einfügen neuer Datensätze jeweils im Anschluss an einen mit „*“ gekennzeichneten Datensatz, Berechnung der absoluten Zeit in Stunden, beginnend mit dem ersten Datensatz eines Patienten, scheint man nur mit auf der Datenmatrix operierenden Programmen bewältigen zu können. Ziel dieses Beitrags soll es sein zu zeigen, dass man im wesentlichen mit DATA-Steps und wenigen SAS-Prozeduren auskommt, ohne auf Programmier-techniken zurückgreifen zu müssen.

Von entscheidender Bedeutung dabei ist die Funktion LAGn. Diese greift auf vorangehende Datensätze zurück¹ Eine ähnliche Funktion sucht man selbst bei komfortablen Datenbanksystemen vergeblich.

Zudem sind fast alle im weiteren benutzten Prozeduren über den SAS-Assistenten erreichbar, gehören damit zum Standardrepertoire eines Anwenders.

3. Die Realisierung

Mit dem nachfolgenden Programmschritt 1 sollen die Behandlungslücken erkannt werden. Die Hilfsvariablen START bzw. ENDE sollen den Start bzw. das Ende der Infusion unter Verknüpfung der Datums-Variablen INFDATS und der Uhrzeit-Variablen INFTIMS bzw. INFDATE und INFTIME beschreiben. Das Datum wird SAS-intern als Integer-Größe gespeichert und gibt die Anzahl der Tage seit dem 1.1.1960 an. Die Uhrzeit wird ebenfalls als Integer-Größe gespeichert und gibt die Anzahl der Sekunden seit 0.00 Uhr an. Deshalb müssen die Datumswerte mit 24 multipliziert und die Uhrzeit durch 3600 dividiert werden, um die Zeitangaben in den Variablen als Stundenangaben vorliegen zu haben. Die Summe dieser beiden Größen gibt die absolut vergangene Zeit seit dem 1.1.1960 um 0.00 Uhr in Stunden an.

Den Wert der Variable LUECKE für einen Datensatz berechnet man als Differenz zwischen Infusionsstart START und dem Infusionsende ENDE des vorhergehenden Datensatzes. Die SAS-Funktion LAGn greift auf den n-ten zurückliegenden Datensatz zu, mit $LAG1(ENDE)=LAG(ENDE)$ hat man folglich das Infusionsende des vorangehenden Datensatzes beschrieben.

Die aus Programmschritt 1 resultierende Datei ist in Tabelle 2 ausgegeben. Die Variable LUECKE beschreibt die Differenz zwischen Infusionsbeginn und dem Ende der Infusion des vorhergehenden Datensatzes mit zwei Besonderheiten:

- Der Wert von LUECKE im ersten Datensatz nimmt den Wert „missing“ an. $LAG(ENDE)$ nimmt an dieser Stelle den Wert „missing“ an, folglich auch die Variable LUECKE.
- Der erste Wert von LUECKE jedes neuen Patienten ist ohne inhaltliche Bedeutung, weil bei der Differenzbildung Werte des vorangegangenen Patienten verwendet wurden. Diese Werte sind mit „missing“ zu belegen. Das wird erst später im Programmschritt 4 realisiert.

¹ Die Zuweisung $VAR_NEU = LAG3(VAR_ALT)$ schreibt auf den n-ten Datensatz für VAR_NEU den Wert des (n-3)-ten Datensatzes der Variable VAR_ALT. Die ersten beiden Datensätze enthalten für VAR_NEU den Wert missing.

Große Aufmerksamkeit sollte bei bedingten Anweisungen der LAG-Funktion geschenkt werden, weil hier einige Besonderheiten zu beachten sind.

Tabelle 2: SAS-Datei ACH.HB2

ISUBJECT	INFDATS	INFTIMS	INFDATE	INFTIME	LUECKE	START
10901	23APR94	2:05	23APR94	11:30	.	300746.08
10901	23APR94	11:30	23APR94	20:30	0.000	300755.50
10901	23APR94	20:30	24APR94	10:45	0.000	300764.50
10901	24APR94	14:30	24APR94	23:00	3.750	300782.50
10901	24APR94	23:00	25APR94	8:30	0.000	300791.00
10901	25APR94	8:30	25APR94	18:00	0.000	300800.50
10901	25APR94	18:00	26APR94	3:30	0.000	300810.00
10901	26APR94	3:30	26APR94	13:00	0.000	300819.50
10901	26APR94	13:00	27APR94	0:30	0.000	300829.00
10901	27APR94	0:30	27APR94	10:00	0.000	300840.50
10901	27APR94	10:00	27APR94	14:30	0.000	300850.00
10901	27APR94	19:30	28APR94	5:00	5.000	300859.50
10901	28APR94	5:00	28APR94	14:30	0.000	300869.00
10901	28APR94	14:30	01MAY94	0:00	0.000	300878.50
10901	01MAY94	0:00	04MAY94	8:00	0.000	300936.00
<hr/>						
11201	19MAY94	12:00	28MAY94	12:10	364.000	301380.00
<hr/>						
11601	03JUN94	22:00	05JUN94	22:00	153.833	301750.00
11601	06JUN94	0:00	06JUN94	2:00	2.000	301800.00
11601	06JUN94	4:00	07JUN94	12:00	2.000	301804.00
11601	07JUN94	14:00	08JUN94	20:00	2.000	301838.00

Vorausgesetzt wird im weiteren, dass die Datei nach den Variablen ISUBJECT, INFDATS und INFTIMS geordnet ist. Im folgenden soll der erste Datensatz jedes Patienten erkannt werden, insbesondere auch, um den Variablenwert von LUECKE auf „missing“ zu setzen. Die Prozedur MEANS erkennt das Minimum der Variablen Start für jeden Patienten, wenn als BY-Variablen ISUBJECT gewählt wird. Die Minima je Patient werden in eine Ausgabedatei OUTMIN gegeben. Der folgende Programmschritt 2 realisiert die in Tabelle 3 ausgegebene OUTMIN-Datei.

Tabelle 3: Datei ACH.OUTMIN

ISUBJECT	_TYPE_	_FREQ_	MIN1
10901	0	15	300746.08
11201	0	1	301380.00
11601	0	4	301750.00

Die Datei OUTMIN enthält neben der Individualkennung ISUBJECT zwei Systemvariable, erkennbar an den Namensbildungen mit „_“. Die Systemmitteilung _FREQ_ beispielsweise gibt die Anzahl der Datensätze an, die in die Minimumbildung eingegangen sind. MIN1 ist die Startzeit der ersten Behandlung des Patienten. Man verknüpft die Datei ACH.HB2 mit ACH.OUTMIN durch die im SAS-Assistenten als kartesisches Produkt bezeichnete MERGE-Prozedur. Ergebnis ist die in Tabelle 4 ausgegebene Datei.

Tabelle 4: Datei ACH.HB3

ISUBJECT	INFDATS	INFTIMS	START	MIN1
10901	23APR94	2:05	300746.08	300746.08
10901	23APR94	11:30	300755.50	300746.08
10901	23APR94	20:30	300764.50	300746.08
10901	24APR94	14:30	300782.50	300746.08
10901	24APR94	23:00	300791.00	300746.08
10901	25APR94	8:30	300800.50	300746.08
10901	25APR94	18:00	300810.00	300746.08
10901	26APR94	3:30	300819.50	300746.08
10901	26APR94	13:00	300829.00	300746.08
10901	27APR94	0:30	300840.50	300746.08
10901	27APR94	10:00	300850.00	300746.08
10901	27APR94	19:30	300859.50	300746.08
10901	28APR94	5:00	300869.00	300746.08
10901	28APR94	14:30	300878.50	300746.08
10901	01MAY94	0:00	300936.00	300746.08
<hr/>				
11201	19MAY94	12:00	301380.00	301380.00
<hr/>				
11601	03JUN94	22:00	301750.00	301750.00
11601	06JUN94	0:00	301800.00	301750.00
11601	06JUN94	4:00	301804.00	301750.00
11601	07JUN94	14:00	301838.00	301750.00

Tabelle 5: Datei ACH.HB4

ISUBJECT	INFDATS	INFTIMS	START	MIN1	DIFFT
10901	23APR94	2:05	300746.08	300746.08	0.000
10901	23APR94	11:30	300755.50	300746.08	9.417
10901	23APR94	20:30	300764.50	300746.08	18.417
10901	24APR94	14:30	300782.50	300746.08	36.417
10901	24APR94	23:00	300791.00	300746.08	44.917
10901	25APR94	8:30	300800.50	300746.08	54.417
10901	25APR94	18:00	300810.00	300746.08	63.917
10901	26APR94	3:30	300819.50	300746.08	73.417
10901	26APR94	13:00	300829.00	300746.08	82.917
10901	27APR94	0:30	300840.50	300746.08	94.417
10901	27APR94	10:00	300850.00	300746.08	103.917
10901	27APR94	19:30	300859.50	300746.08	113.417
10901	28APR94	5:00	300869.00	300746.08	122.917
10901	28APR94	14:30	300878.50	300746.08	132.417
10901	01MAY94	0:00	300936.00	300746.08	189.917
<hr/>					
11201	19MAY94	12:00	301380.00	301380.00	0.000
<hr/>					
11601	03JUN94	22:00	301750.00	301750.00	0.000
11601	06JUN94	0:00	301800.00	301750.00	50.000
11601	06JUN94	4:00	301804.00	301750.00	54.000
11601	07JUN94	14:00	301838.00	301750.00	88.000

Diejenigen Datensätze, bei denen START und MIN1 übereinstimmen, sind die ersten Daten-

sätze eines Patienten. Für genau diese wird die Zeitdifferenz $\text{DIFFT}=\text{START}-\text{MIN1}$ gleich Null und für die Variable LUECKE ist auf „missing“ zu setzen (siehe Programm 4 und Tabelle 5).

In die Datei HB4 werden zwei Hilfsvariablen a und b eingefügt, auf die das Infusionsende (INFDATS und INFTIMS) zwischengespeichert wird. Aus dieser Datei werden diejenigen Datensätze in ACH.HB5 kopiert, bei denen die Variable LUECKE positive Werte annimmt. Eine Umspeicherung der Variablen in der Datei ACH.HB5 kann erfolgen: Auf das Infusionsende kann der Infusionsanfang gespeichert werden, auf den Infusionsanfang die zwischengespeicherten Werte a und b (Ende des vorhergehenden Infusionsintervalls= Anfang des „therapiefreien“ Intervalls). Zusätzlich müssen die übrigen in der Datei enthaltenen Variablen (START, ENDE, DIFFT, LUECKE, INFSPEED) für den Datensatz mit den neuen Start- und Endzeiten neu berechnet werden.

Die Datei ACH.HB6 wird an die Datei ACH.HB4 angehängt, anschließend nach den Variablen ISUBJECT, INFDATE und INFTIME sortiert. Damit steht sie als Eingangsdatei ACH.HB7 (siehe Tabelle 7) für die Prozedur REG zur Verfügung. Diese kann mit der unabhängigen Variable DIFFT und der abhängigen INFSPEED für jeden Patienten (BY-Variablen ISUBJECT) abgearbeitet und die Schätzungen für den Anstieg und das absolute Glied der Regressionsgeraden in der SAS-Datei ACH.OUTREG abgelegt werden. Die Anstiege geben für jedes Individuum Auskunft darüber, ob die Dosierung global, d.h. über die gesamte Behandlungszeit gesehen, angestiegen oder abgefallen ist.

Tabelle 6: Datei ACH.HB7, mit der die Regressionsanalyse durchgeführt werden kann

ISUBJECT	INFDATE	INFTIME	INFDATE	INFTIME	DIFFT	INFSPEED
10901	23APR94	2:05	23APR94	11:30	0.000	5.6
10901	23APR94	11:30	23APR94	20:30	9.417	5.6
10901	23APR94	20:30	24APR94	10:45	18.417	3.8
10901	24APR94	10:45	24APR94	14:30	32.667	0.0
10901	24APR94	14:30	24APR94	23:00	36.417	5.6
10901	24APR94	23:00	25APR94	8:30	44.917	5.6
10901	25APR94	8:30	25APR94	18:00	54.417	5.6
10901	25APR94	18:00	26APR94	3:30	63.917	5.6
10901	26APR94	3:30	26APR94	13:00	73.417	5.6
10901	26APR94	13:00	27APR94	0:30	82.917	5.6
10901	27APR94	0:30	27APR94	10:00	94.417	5.6
10901	27APR94	10:00	27APR94	14:30	103.917	5.6
10901	27APR94	14:30	27APR94	19:30	108.417	0.0
10901	27APR94	19:30	28APR94	5:00	113.417	5.6
10901	28APR94	5:00	28APR94	14:30	122.917	5.6
10901	28APR94	14:30	01MAY94	0:00	132.417	5.6
10901	01MAY94	0:00	04MAY94	8:00	189.917	5.6
11201	19MAY94	12:00	28MAY94	12:10	0.000	4.9
11601	03JUN94	22:00	05JUN94	22:00	0.000	2.8
11601	05JUN94	22:00	06JUN94	0:00	48.000	0.0
11601	06JUN94	0:00	06JUN94	2:00	50.000	2.2
11601	06JUN94	2:00	06JUN94	4:00	52.000	0.0
11601	06JUN94	4:00	07JUN94	12:00	54.000	1.8
11601	07JUN94	12:00	07JUN94	14:00	86.000	0.0
11601	07JUN94	14:00	08JUN94	20:00	88.000	1.4

Die Ergebnisse der Regressionsanalyse für die drei Patienten sind wie folgt:

- Pat. 10901 hat einen sehr kleinen, aber positiven Anstieg(+0.003876). Die Dosierung hat sich im Laufe der Behandlung im wesentlichen erhöht.
- Pat. 11201 hat den Anstieg Null. Er ist durchgängig mit einer Dosierung behandelt. Die Prozedur REG berechnet bei Vorliegen nur eines Meßwertes als Regressionsgerade eine konstante Funktion. Das absolute Glied stimmt mit der Dosierung überein.
- Pat. 11601 hat einen negativen Anstieg, seine Dosierung ist für die Behandlungsintervalle im wesentlichen gesenkt worden.

Tabelle 7: Ergebnisse der Regressionsanalyse (DIFFT=Anstieg der Regressionsgeraden, Intercept=Absolutes Glied der Regressionsgeraden)

ISUBJECT	Intercept	DIFFT
10901	4.54303	0.003876
11201	4.90000	0.000000
11601	2.34953	-0.021817

4. Kritische Schlußbemerkungen

Die Modellvorstellung selbst, dass der Anstieg der Regressionsgeraden Auskunft darüber gibt, ob die Infusionsgeschwindigkeit global über die gesamte Behandlungszeit des Patienten hin im wesentlichen erhöht oder erniedrigt wurde, ist nur bedingt richtig. Der positive Anstieg des Patienten 10901 resultiert allein aus den Anordnungen der Behandlungslücken und der einmaligen Reduktion der Infusionsgeschwindigkeit (erster Datensatz in Tabelle 6). Auch der Ausweg, besonders kleine Anstiege der Regressionsgeraden als Null anzusehen, bringt keine Abhilfe. Vielmehr würde man den Zeitverlauf der Infusionsgeschwindigkeiten eher als eine Reduktion ansehen.

Besser wäre der Ansatz, die Ausgleichsgerade für die Infusionsgeschwindigkeit am Startpunkt festzumachen. Dafür kann mit Erfolg die Option „noint“- kein absolutes Glied der Regressionsgeraden (intercept) - angewandt werden. Dazu sind die folgenden Schritte nötig:

- Die Startinfusionsgeschwindigkeit jedes Patienten wird von sämtlichen Infusionsgeschwindigkeiten subtrahiert. Dadurch erreicht man, dass zum Zeitpunkt Null die Infusionsgeschwindigkeitsdifferenz Null wird.
- Die Prozedur REG mit der Option NOINT wird auf die Infusionsgeschwindigkeitsdifferenzen als abhängige Variable und die absolute Zeit DIFFT als unabhängige Variable angewandt bezüglich der BY-Variable ISUBJECT.

Die resultierenden Anstiege entsprechen den Anstiegen der Regressionsgerade, die am Startwert festgemacht wurde.

Tabelle 8: Ergebnisse der Regressionsanalyse (DIFFT=Anstieg der Regressionsgeraden)

ISUBJECT	DIFFT
10901	-0.006000
11201	0.000000
11601	-0.021817

5. Ergebnisse

Von den 213 Teilnehmern der Studie hatte bei 27 Patienten die Regressionsgerade den Anstieg Null. Diese sind von Therapiebeginn bis zum Ende der Behandlung mit der Eingangsdosis ausgekommen. Der überwiegende Anteil (n=152) aber hatte einen negativen Anstieg der Regressionsgeraden. Die Dosen mußten im wesentlichen reduziert werden. Damit ist eine Hauptaussage der Studie umrissen: Die Eingangsdosis wurde zu hoch angesetzt.

6. SAS-Programm

```
/*-----*
| Programmschritt 1 zum Erkennen von Behandlungsluecken |
| Eingabedatei: ACH.HB1 (siehe Tabelle 1) |
| Ausgabedatei: ACH.HB2 (siehe Tabelle 2) |
*-----*/
data ACH.HB2;
  set ACH.HB1;
  START = INFDATS*24+INFTIMS/3600;
  ENDE = INFDATE*24+INFTIME/3600;
  LUECKE=START-LAG(ENDE);
run;

/*-----*
| Programmschritt 2 zum Erkennen des ersten Datensatzes |
| jedes Patienten, entspricht kleinstem Wert der |
| Variablen START eines jeden Patienten |
| Eingabedatei: ACH.HB2 (siehe Tabelle 2) |
| Ausgabedatei: ACH.OUTMIN (siehe Tabelle 3) |
*-----*/
proc means data=ACH.HB2
  MIN;
  var START ;
output out=ACH.OUTMIN
  MIN=MIN1;
  by ISUBJECT;
run;
```

```

/*-----*
| Programmschritt 3: Verknuepfen der Datei ACH.HB2 mit |
| ACH.OUTMIN (kartesisches Produkt) |
| Eingabedateien: ACH.HB2 und ACH.OUTMIN |
| Ausgabedatei: ACH.HB3 (siehe Tabelle 4)2 |
*-----*/
data ACH.HB3;
  merge ACH.HB2 ACH.OUTMIN;
  by ISUBJECT;
run;

/*-----*
| Programmschritt 4: Berechnung der absoluten Zeit DIFFT |
| Eingabedatei: ACH.HB3 (siehe Tabelle 4) |
| Ausgabedatei: ACH.HB4 (siehe Tabelle 5) |
*-----*/

data ACH.HB4;
  set ACH.HB3;
DIFFT = START-MIN1;
  if DIFFT=0 then LUECKE= .;
run;

/*-----*
| Programmschritt 5: An ACH.HB4 werden die Variablen |
| a und b angehaengt, danach Unterdatei ACH.HB5 erzeugt |
| (ausschliesslich Datensaeetze mit echter LUECKE) |
*-----*/

data ACH.HB4;
  set ACH.HB4;
a=LAG(INFDATE);
b=LAG(INFTIME);
run;
data ACH.HB5;
  set ACH.HB4;
  where LUECKE ^= 0 AND LUECKE ^= .;
run;

```

² Wenn die Verknüpfung über den SAS-Assistenten erfolgt, wird die folgende kompliziertere Programmstruktur gewählt:

```

proc sql;
  create table ACH.HB3 as select *
  from ACH.HB2 ,ACH.OUTMIN
  where HB2.ISUBJECT = OUTMIN.ISUBJECT ;
quit;
run;

```

```
/*-----*
| Programmschritt 6: Umspeichern der Datums- und      |
| Zeitangaben und Neuberechnung von Variablen        |
| Eingabedatei: ACH.HB5                              |
| Ausgabedatei: ACH.HB6                              |
*-----*/

data ACH.HB6;
    set ACH.HB5;
INFDATE = INFDATS;
INFTIME = INFTIMS;
INFDATS = A;
INFTIMS = B;
START=INFDATS*24+INFTIMS/3600;
ENDE=INFDATE*24+INFTIME/3600;
DIFFT=START-MIN1;
LUECKE = 0;
INFSPEED = 0;
run;

/*-----*
| Programmschritt 7: Anhaengen der Datensaeetze HB6 an HB4 |
| und Einsortierung nach den Startzeiten der Infusion pro |
| Patient                                                  |
| Eingabedateien: ACH.HB4 ACH.HB6                        |
| Ausgabedatei: ACH.HB7                                  |
*-----*/

data ACH.HB7;
    set ACH.HB4 ACH.HB6;
run;
proc sort data = ACH.HB7;
    by ISUBJECT INFDATS INFTIMS;
run;

/*-----*
| Programmschritt 8: Regressionsanalyse mit INFSPEED als |
| abhaengige und DIFFT als unabhaengige Variable        |
| mit Ergebnisausgabe in Outputdatei ACH.OUTREG        |
| Eingabedatei: ACH.HB7                                  |
| Ausgabedatei: ACH.OUTREG (siehe Tabelle 7)            |
| Von mehreren Output-Moeglichkeiten ist die Ausgabe der |
| Schaetzungen (estimation) der Geradenparameter gewaehlt |
*-----*/

proc reg data=ACH.HB7
    outest=ACH.OUTREG;
    by ISUBJECT;
    model INFSPEED= DIFFT;
run;
quit;
```

7. Literaturhinweis

SAS Institute Inc., SAS Language: Reference, Version 6, First Edition Cary, NC: SAS Institute Inc., 1990, 1042 pp.

Bernd Jäger
Ernst-Moritz-Arndt-Universität
Institut für Biometrie und Medizinische Informatik
Klinikum Sauerbruchstraße
17487 Greifswald
jaeger@biometrie.uni-greifswald.de