

Auswertung von Signalintensitätsdaten bei MRI-Mammographien mit Kontrastmitteln

Dr. H. Wöhling, Nycomed Arzneimittel GmbH, Ismaning bei München

Eine klinische Studie

Eine offene, randomisierte, klinische Studie der Phase II diente der Wahl der optimalen aus 3 möglichen Dosen eines Kontrastmittels bei kernspintomographischen (MRI) Mammographien. Es handelte sich um eine Studie im Parallelgruppendesign, in die 18 Patienten mit vermutlich gutartigen (benignen; engl. Benign) Brusttumoren und 18 Patienten mit vermutlich bösartigen (malignen; engl. Malignant) Brusttumoren eingeschlossen wurden. Alle Frauen wurden kernspintomographisch mit Gabe eines Kontrastmittels untersucht. Es wurden dynamische Sequenzen zur Untersuchung benutzt, die Signalintensitätsmessungen von vor der Kontrastmittelgabe bis 9 Minuten nach Kontrastmittelgabe zuließen. Anhand der resultierenden Signalintensitätskurven, die die Kontrastmittelaufnahme der Tumoren charakterisieren, sollten sogenannte Cut-off Werte definiert werden, um zwischen benignen und malignen Tumoren unterscheiden zu können. Je nachdem, wie gut anhand der definierten Cut-off Werte eine Unterscheidung möglich wurde, sollte auf eine optimale Dosis geschlossen werden.

Die Auswertungsstrategie

Bisherige Erfahrungen mit derartigen Signalintensitätskurven (SI-Kurven) weisen darauf hin, daß für maligne Tumoren im allgemeinen ein initial sehr steiler und großer Signalanstieg mit einem Plateau nach 2-3 Minuten charakteristisch ist, während benigne Tumore im allgemeinen kaum einen oder einen kontinuierlicher ansteigenden Signalverlauf zeigen. Diese Charakteristika galt es mit statistischen Methoden zu unterscheiden.

Dazu wurden die originalen Signalintensitätskurven zunächst geglättet und dann standardisiert. Nach der Definition von Cut-off Kriterien zur Unterscheidung von bösartigen und gutartigen Tumoren wurden Accuracy, Sensitivität und Spezifität berechnet, wobei die MRI-Diagnose mit dem Ergebnis der Histologie verglichen wurde.

Glättung der Signalintensitätskurven

Die Glättung der Kurven erfolgte nach dem Prinzip der gleitenden Mittelwerte über drei aufeinanderfolgende Werte. Ausgehend von dem originalen Datensatz der Form

PATNO	REG_NUM	IMAGE_NO	SI_MEAN
1	2	570	264
1	2	571	274
1	2	572	266
1	2	573	260
1	2	574	266
1	2	575	268
.	.	.	.
.	.	.	.
.	.	.	.

wurde im SAS-Job ein PROC TRANSPOSE benutzt, bevor in einem DATA step das eigentliche Smoothing programmiert wurde. Nach einem entsprechenden PROC TRANSPOSE für die Rücktransformation der geglätteten Werte wurden in anschließenden DATA steps die Zeitachse konstruiert.

Standardisierung der Kurven

Die absoluten Signalintensitätskurven sollten gegen eine bzw. mehrere Standards abgeglichen werden, um sie gegeneinander bezüglich der interessierenden Aspekte vergleichbar zu machen. Zur Verfügung standen in einigen Fällen Meßkurven eines benutzten (maschinenspezifischen) Phantoms, SI-Messungen des umgebenden Fett- und Muskelgewebes. Die Standardisierung erfolgte jeweils nach der Formel

$$SI_{st,i,k}(t) = (SI_k(t) - \text{mean}(SI_{i,k})) / SD_{i,k} \quad \text{mit } i = 1,2,3 ; k = 1, \dots, 36$$

wobei i die Art der Standardisierung kennzeichnet und k die Patienten markiert.

In SAS wurde die Standardisierung mittels eines PROC MEANS programmiert, welches ein BY-Statement bezüglich der Patienten und der Meßregion enthielt und als Output einen Dataset produzierte, der Mittelwerte und Standardabweichungen enthielt. Einfache anschließende Datasteps komplettierten die Standardisierung. Anhand zweier Beispiele, einem gutartigen und einem bösartigen Tumor, mit typischen Kurvenverläufen und Standardmessungen kann die Reliabilität der verschiedenen Standardisierungen verdeutlicht werden:

Tumorart	standardisiert nach	mean(SI)	SD
benigne	Phantom	262	4.4
	Fett	208	9.7
	Muskel	127	12.9
maligne	Phantom	219	4.3
	Fett	195	5.5
	Muskel	105	8.3

Die maschinenspezifischen Phantommessungen haben die geringste Variabilität aber einen stärkeren Kontrast als Fett- und Muskelgewebe. Das Muskelgewebe zeigt die größte Variabilität, da es leicht kontrastiert. Bei einer Standardisierung der Tumor-SI-Kurven zum Muskelgewebe geht daher viel an Variabilität verloren und eine Unterscheidung der Tumorarten ist kaum noch möglich. Eine Standardisierung zu einem maschinenspezifischen Phantom ist am ehesten sinnvoll, da damit nur die Variabilität der Maschine eliminiert wird.

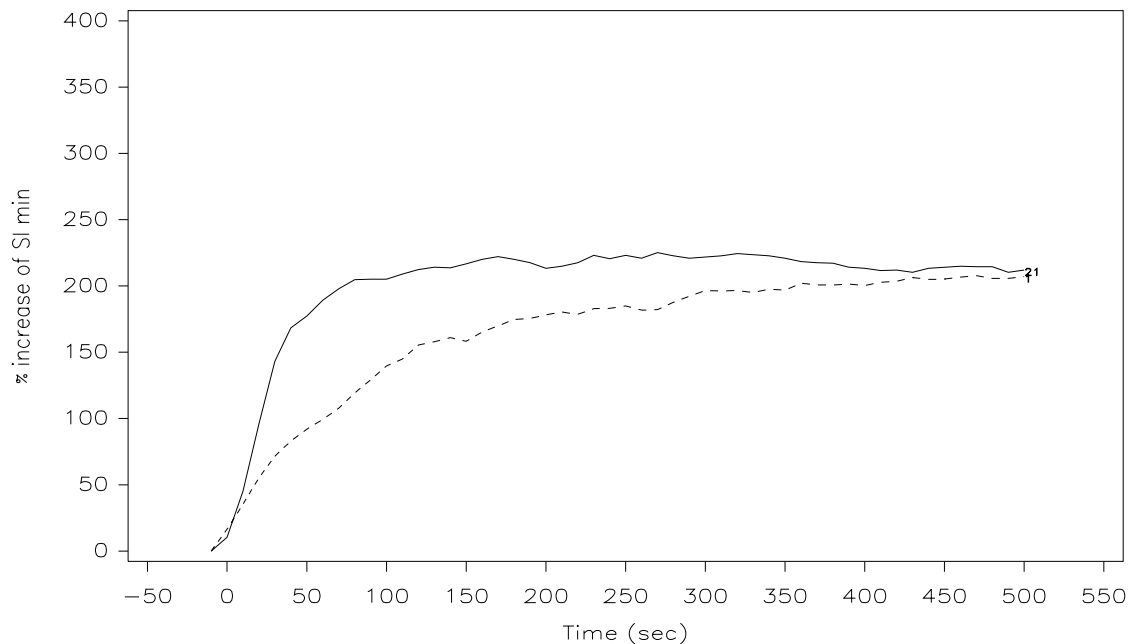
Nach jeder dieser Standardisierungen bleibt das Problem, daß sämtliche SI-Kurven auf unterschiedlichen Levels liegen und die absoluten SI-Einheiten nicht interpretierbare Größen sind. Eine Vergleichbarkeit der Kurven ist somit nur sehr bedingt möglich.

Daher wurde desweiteren aus jeder absoluten SI-Kurve eine relative (prozentuale) Kurve im Verhältnis zum Wert vor Kontrastmittelaufnahme berechnet. Dies geschah gemäß der Formel

$$SI_{\%,k}(t) = (SI_k(t) - SI_{pre,k}) * 100 / SI_{pre,k} \quad \text{mit } k = 1, \dots, 36$$

wobei k die Patienten markiert.

Die so standardisierten Kurven sind direkt vergleichbar und die prozentualen Anstiegsunterschiede direkt im Verhältnis zu den Prekontrastwerten interpretierbar. Die folgende Grafik zeigt charakteristische, prozentual standardisierte Kurvenverläufe eines benignen und eines malignen Tumors.



Legend: continuous line = malignant lesion
dotted line = benign lesion

Damit eignen sich die so standardisierten Kurven als einzige für eine sinnvolle Definition des gesuchten Kriteriums.

Definition von Cut-off Kriterien

Für die Suche nach einem geeigneten Cut-off Kriterium standen verschiedene Kurvenparameter zur Verfügung:

- maximale Signalintensität über die gesamte Kurve
- maximale Signalintensität innerhalb der ersten 2-3 Minuten nach Signalanstieg
- Slope des steilsten Anstiegs

Die maximalen Signalintensitäten erwiesen sich als unbrauchbar. Für ein Slope-Kriterium mußte zunächst der steilste Anstieg definiert werden. Es erwies sich als geeignet, den ersten Kurvenabschnitt auszuwählen, in dem die aufeinanderfolgenden Meßpunkte eine mindestens 10% ige Signalverstärkung aufwiesen. Sollte dies in den ersten 2 Minuten nach Beginn des Signalanstiegs nie der Fall gewesen sein, so wurden alle Meßpunkte der ersten 2 Minuten nach beginnendem Signalanstieg gewählt. In einer linearen Regression über die ausgewählten Meßpunkte wurden

dann die gewünschten Slopes ermittelt. Im SAS-Job war dazu eine Macro-Programmierung mit eingeschlossener Regressions-analyse notwendig.

```
%MACRO pat (p);

DATA reg&p;
  SET regres;
  WHERE (patno = &p and reg_num=1);
RUN;

TITLE "Patient &p: LINEAR REGRESSION OF THE CM UPTAKE y = a + bx";
PROC NLIN DATA = reg&p
  METHOD = MARQUARDT
  CONVERGEPARM = 0.00001
  BEST = 10;
  PARMS a = -200 TO 100 BY 10
        b = 0 TO 15 BY 1;
  MODEL si_rel = a + b*time;
  der.a = 1;
  der.b = time;
  OUTPUT OUT = fit&p parms = a parms = b;
RUN;
QUIT;
%MEND;

%pat (1);
%pat (2);
.....
```

Das Ergebnis der Berechnungen stellt die folgende Tabelle dar.

Tumor	Dosis	Slope				
		mean	std	min	max	n
benigne	hoch	1.36	1.48	0.33	3.47	4
	mittel	1.19	1.77	0.09	4.78	7
	gering	0.42	0.44	0.08	1.18	5
maligne	hoch	2.44	0.65	1.70	3.40	5
	mittel	2.22	0.81	1.39	3.02	3
	gering	2.97	5.07	0.06	11.96	5

Wider Erwarten ergab sich auch für den Slope kein reliables Kriterium.

Letztendlich erwiesen sich dosisabhängige Schwellenwerte 2 Minuten nach Beginn des Signalanstiegs von

- 50% relativer Signalanstieg für die geringste Dosis
- 70% relativer Signalanstieg für die mittlere Dosis
- 90% relativer Signalanstieg für die höchste Dosis

als zuverlässigstes Kriterium. Dabei wurde ein Tumor als maligne charakterisiert, wenn der relative Signalanstieg nach 2 Minuten oberhalb des Schwellenwertes lag und entsprechend als benigne gekennzeichnet, wenn der Signalanstieg geringer war.

Studienergebnis

Berechnungen von Accuracy, Senitivität und Spezifität zeigten folgendes Ergebnis:

Dosis	Accuracy	Sensitivität	Spezifität
hoch	0.70	0.60	0.80
mittel	0.70	0.67	0.71
gering	0.78	1	0.50

Das Ergebnis bezüglich Accuracy, Sensitivität und Spezifität zeigt eine Tendenz zu höherer Sensitivität bei geringerer Spezifität mit höherer Dosis. Da hinter der Sensitivität die Charakterisierung maligner Tumore steht, ist vom medizinischen Standpunkt eine höhere Dosis empfehlenswert.

Dieses Ergebnis, insbesondere die Validität der Cut-off Werte, müßte nun an einer weiteren Studie geprüft werden.

Kernelemente der SAS Programmierung

In diesem Beispiel aus der medizinischen Forschung basierte die SAS Programmierung primär auf den vielfältigen Möglichkeiten der DATA step- und MACRO-Programmierung.

Die benötigten Prozeduren waren:

- PROC TRANSPOSE beim Data Handling
- PROC MEANS für die Standardisierungen
- PROC NLIN für die Parameterberechnungen
- PROC TABULATE für die Ergebnisdarstellung
- PROC GPLOT für die Visualisierung der Daten

Wenn auch die einzelnen Programmteilen keine anspruchsvollen Programmierungen benötigten, so ist dies insgesamt ein komplexes Anwendungsbeispiel einer SAS Programmierung aus der praktischen medizinischen Forschung.