



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Pruscha, Ulm, Schmidt:

## Statistische Analyse des Einflusses von Herzrhythmusstörungen auf das Mortalitätsrisiko

Sonderforschungsbereich 386, Paper 56 (1997)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



# Statistische Analyse des Einflusses von Herzrhythmusstörungen auf das Mortalitätsrisiko

Helmut Pruscha, Kurt Ulm und Georg Schmidt

Mathematisches Institut der LMU München  
Institut für Medizinische Statistik der TU München  
I. Medizinische Klinik der TU München

## 1 Einleitung

Herzrhythmusstörungen stellen eine äußerst bedrohliche Krankheit dar und können zum plötzlichen Herztod führen. So sterben in der Bundesrepublik Deutschland pro Jahr etwa 100 000 Patienten an einem Herz-Kreislauf-Stillstand, der in 65 - 80 % durch eine Rhythmusstörung hervorgerufen wird (Trappe et al., 1996). Seit über 20 Jahren ist bekannt, daß das Ausmaß der Rhythmusstörungen wesentlich das Risiko für einen plötzlichen Herztod beeinflußt (Moss et al., 1979).

Die Identifizierung von Patienten mit einem erhöhten Mortalitätsrisiko ist daher von erheblichem Interesse und nach wie vor noch nicht zufriedenstellend gelöst. Von dieser Frage hängt die Wahl der geeigneten Therapie ab. Bei Patienten mit einem erhöhten Mortalitätsrisiko ist derzeit die Implementierung eines Defibrillators die einzig wirksame Therapie. Die medikamentöse Behandlung mit sog. Antiarrhythmika war lange Zeit die Therapie der Wahl, bis Ende der 80er Jahre eine Studie aus den USA für einige Medikamente ein erhöhtes Mortalitätsrisiko nachwies (CAST-Studie, 1989).

Seit dieser Zeit konzentriert sich die Forschung auf zwei Punkte, die Entwicklung neuer Medikamente und die Erkennung von besonders gefährdeten Patienten. Die einzige nicht-invasive Methode zur Erfassung der Häufigkeit

der Arrhythmien ist gegenwärtig das 24 Std. Holter-EKG. Derzeit wird für die Unterteilung in verschiedene Risikogruppen nur das Ausmaß der Rhythmusstörungen, die Häufigkeit der sog. ventrikulären Extrasystolen (VES) erfaßt. Dieser Faktor ist aber nicht aussagekräftig genug. Daher liegt es nahe, die Information über die Rhythmusstörungen besser zu nutzen und vor allem die Komplexität der Arrhythmien besser zu beschreiben. Hierzu werden aus dem 24 Std. Holter-EKG alle Abstände zwischen zwei aufeinanderfolgenden Herzschlägen, die sog. RR-Intervalle, erfaßt. Wenn im Durchschnitt ein Herzschlag pro Sekunde erfolgt, liegen über 24 Stunden ca. 90 000 solcher Intervalle vor. Diese Datenmenge stellt an die Analyseverfahren eine besondere Herausforderung dar.

In einem ersten Ansatz wurden Methoden aus dem Bereich der nichtlinearen Dynamik angewandt (Schmidt et al., 1996). Es ist bekannt, daß neben den Rhythmusstörungen auch die Variabilität der RR-Intervalle das Risiko beeinflussen. Mit den Ansätzen, basierend auf der nichtlinearen Dynamik, wurden aus den Daten eines 24 Std. Holter-EKG's zwei Parameter abgeleitet ( $\alpha_{VES}$  und  $\alpha_{sin}$ ). Der erste Parameter beschreibt die Komplexität, der zweite steht für die Variabilität.

Die vorliegende Arbeit wendet statistische Verfahren aus den Bereichen

Kurvenschätzung, logistische Regression, Coxsche Regression

an, um besonders gefährdete Patienten zu erkennen. Für diese Analyse standen die Daten von 60 Patienten zur Verfügung. Das Ziel dieser Untersuchung ist es insbesondere, die aufwendige Methode der Bestimmung von  $\alpha_{sin}$ ,  $\alpha_{VES}$  durch eine neue zu ersetzen, die konzeptionell und numerisch einfacher ist, die – im Unterschied zur eingeführten – vollständig algorithmisch durchgeführt werden kann und die auch – bei entsprechender Weiterentwicklung – zum Teil online erfolgen könnte.

## **2 Ermittlungen der Prädiktoren**

### **2.1 Histogramm der Originaldaten**

Das Langzeit-EKG eines einzelnen Patienten besteht aus einer Folge

$$x_1, x_2, \dots, x_n \tag{1}$$

von aufeinanderfolgenden Intervall-Längen. Der Wert  $x_i$  gibt die Zeitdauer zwischen dem  $i$ -ten und  $(i + 1)$ -ten Herzschlag an, gemessen in  $1/1000$  sec. Die Anzahl  $n$  liegt dabei zwischen 60 000 und 150 000. Die Verteilung der Werte  $x_i$  wird mit Hilfe von Dichte-Kernschätzern  $\hat{f}_n(x)$  angegeben. Mit einer Fensterbreite  $h = 0.1$  [sec] und mit dem Dreiecksfenster  $K(x) = 1 - |x|$ , falls  $-1 \leq x \leq 1$ , ( $K(x) = 0$  sonst) berechnet man

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \tag{2}$$

und trägt (2) an den Stellen

$$x = (j - 1) * \tau + \frac{1}{2}\tau, \quad j = 1, \dots, J \tag{3}$$

auf (Silverman, 1986). Mit der Wahl von  $\tau = 0.05$  [sec] und  $J = 40$  ergeben sich geglättete Histogramme wie in Abb. 1.

Zu aussagekräftigen Prädiktoren gelangt man aber nicht durch Histogramme der Intervall-Längen  $x_i$  aus (1), sondern durch Histogramme von geeigneten Transformationen.

## 2.2 Histogramm der transformierten Daten

Nach Vorgabe einer ungeraden Zahl  $m = 2k + 1$  bilden wir für jedes Intervall  $x_i$  (genauer: Intervall-Länge  $x_i$ ) aus der Folge (1) den Block

$$x_{i-k}, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{i+k} \tag{4}$$

von  $k$  aufeinanderfolgenden Intervallen, mit  $x_i$  als dem mittleren Intervall. Der Block (4) stellt die 'Umgebung' des Zeitintervalls  $x_i$  dar, der Mittelwert dieser  $m$  Zahlen, also

$$\bar{x}_i = \frac{1}{m} \sum_{j=-k}^k x_{i+j}, \quad (5)$$

definiert die für  $x_i$  gültige 'Zeiteinheit', und

$$t_i = \frac{x_i}{\bar{x}_i} \quad (6)$$

schließlich stellt die Länge des  $i$ -ten Intervalls in der neuen Zeiteinheit dar. Anstelle von (1) werten wir nun die Folge

$$t_{k+1}, \dots, t_{n-k} \quad (7)$$

von  $N = n - 2k$  transformierten Intervall-Längen aus (wegen der Operation (5) verliert man am Anfang und am Ende der Folge je  $k$  Stück). Aus (7) berechnet man wie in (2) und (3) die Kernschätzungen

$$\hat{f}_N(t) = \frac{1}{Nh} \sum_{i=k+1}^{n-k} K\left(\frac{t - t_i}{h}\right), \quad (8)$$

mit den dort angegebenen Größen  $K$ ,  $h$ ,  $\tau$ ,  $t$  (anstelle von  $x$ ) und  $J$ . Für zwei ausgewählte Patienten sind entsprechende Histogramme mit den Zahlen  $m = 3, 5$  und  $7$  in Abb. 2 und 3 dargestellt. Zusätzlich sind noch Histogramme eingetragen, bei denen  $\bar{x}_i$  nicht gemäß Formel (5), sondern gemäß  $\sum_{j=0}^{m-1} x_{i-j}/m$  berechnet wurde; bei denen der Block der  $m = 5$  Intervalle also nicht um  $x_i$  herum gelegt wurde, wie in (4) oben, sondern linksseitig von  $x_i$ , d.h. in der Form  $x_{i-m+1}, \dots, x_{i-1}, x_i$ . Zur Identifizierung eines unteren und oberen Flügels (vgl. 2.3 unten) erwiesen sich Histogramme mit  $m = 5$  und mit Blöcken der Form (4) am geeignetsten. Für die nach Formel (8) berechneten Größen  $\hat{f}_N(t)$ ,  $t = (j - 1) * \tau + \frac{1}{2}\tau$ ,  $j = 1, \dots, J$  [ $J = 40$ ,  $\tau = 0.05$ ], schreiben wir kurz

$$H(t), \quad 0 \leq t \leq 2.$$

### 2.3 Zentrum und Flügel des Histogramms

In der Mehrzahl der Fälle weist der Kurvenverlauf von  $H(t)$  unterhalb und oberhalb von  $t = 1$  je ein relatives Minimum auf. Bezeichnen wir die Werte, an denen diese Minima vorliegen, mit  $x_u$  und  $x_o$ , so unterscheiden wir einen unteren Flügel, einen Zentralteil und einen oberen Flügel des Histogramms (Abb. 4a). Entsprechend berechnen wir die folgenden Momente des Histogramms, durch welche sowohl die Masse dieser Histogrammteile als auch die mittlere Entfernung ihrer Punkte vom Zentrum  $t = 1$  berücksichtigt werden:

$$\begin{aligned} \text{unteres Moment } m_u &= \sum_{i \in U} |1 - t_i|/N \\ \text{oberes Moment } m_o &= \sum_{i \in O} |1 - t_i|/N \\ \text{zentrales Moment } m_z &= \sum_{i \in Z} |1 - t_i|/N \end{aligned}$$

wobei  $U$ ,  $O$ ,  $Z$  die Menge der Indizes  $i$  umfaßt, für welche jeweils

$$t_i \leq x_u, t_i \geq x_o \text{ bzw. } x_u < t_i < x_o$$

gilt, und wobei im folgenden  $|U|, |O|, |Z|$  jeweils die Anzahl dieser Indizes angibt ( $|U| + |O| + |Z| = N$ ). Ferner bezeichne  $s_z$  die Standardabweichung aller derjenigen (Original-Intervallängen)  $x_i$ , deren transformierte Länge  $t_i = x_i/\bar{x}_i$  ins Zentrum des Histogramms fällt:

$$s_z^2 = \frac{1}{|Z|} \sum_{i \in Z} (x_i - \bar{x}_z)^2, \quad \bar{x}_z = \frac{1}{|Z|} \sum_{i \in Z} x_i.$$

Ganz analog bezeichnet  $s_f$  die Standardabweichung derjenigen  $x_i$ , deren transformierte Länge  $t_i$  in einen (der beiden) Flügel des Histogramms fällt.

Eine genauere algorithmische Definition der Cutpoint-Werte  $x_u$  und  $x_o$  umfaßt auch Fälle, in denen kein relatives Minimum von  $H(t)$ ,  $0 \leq t \leq 2$ , vorhanden ist (wie in Abb. 4b). Mit einer vorgegebenen Zahl  $\epsilon$  bestimmen wir  $x_o$  als diejenige Stelle (Krümmungspunkt), für welche die aufeinanderfolgenden Differenzen  $\Delta_- = H(x_o - \tau) - H(x_o)$ ,  $\Delta_+ = H(x_o) - H(x_o + \tau)$  die Ungleichung  $\Delta_+/\Delta_- < \epsilon$  erfüllen.

## 3 Statistische Auswertungen

### 3.1 Zugrundeliegende Dateien

Das Langzeit-EKG von 60 Patienten wurde nach der oben beschriebenen Methode analysiert und jeweils auf die folgenden 7 Kenngrößen reduziert:

$$\begin{aligned} x_u, x_o &: && \text{unterer und oberer Cutpoint des Histogramms} \\ m_u, m_o, m_z &: && \text{unteres, oberes und zentrales Moment} \\ s_z, s_f &: && \text{Standardabweichung (der Originaldaten)} \\ &&& \text{im Zentrum und in den Flügeln} \end{aligned} \tag{9}$$

Zusätzlich liegen von jedem Patienten vor:

$$\begin{aligned} Oxf & \text{ Oxford-Kürzel (Kennzeichnung des Patienten)} \\ Status & \text{ 1 (überlebend), 2 (verstorben) oder 3 (PHT)} \\ Zeit & \text{ Zeitdauer von der Aufnahme des EKG bis zum} \\ & \text{ Tod (Status = 2, 3) oder bis zum Verlassen} \\ & \text{ der Studie (Status = 1)} \\ VESh & \text{ Anzahl der pro Stunde aufgetretenen ventrikulären} \\ & \text{ Extrasystolen (VES)} \end{aligned} \tag{10}$$

Der  $Status = 3$  wurde vergeben, falls der Patient (bekanntermaßen) an einem plötzlichen Herztod (PHT) verstorben war. Für alle anderen Todesursachen wurde  $Status = 2$  vorgesehen.

Für die Auswertung wurden von jedem Patienten die 7 neu entwickelten Variablen (9) verwendet (vgl. Tab. 1) sowie die Variable  $VESh$ . Zur Beurteilung der Güte dieses neuen Ansatzes wurde sie verglichen mit Ergebnissen, welche auf der nichtlinearen Dynamik basieren, also auf den Variablen  $VESh$ ,  $\alpha_{sin}$  und  $\alpha_{VES}$  (vgl. Tab. 2). Dabei ist zu bemerken, daß die Berechnung von  $\alpha_{sin}$  und  $\alpha_{VES}$  bei jedem einzelnen Zeitintervall  $x_i$  aus der Folge (1) auf die Information „VES oder kein VES“ zurückgreift, während die Bestimmung der 7 Variablen (9) ohne diese Zusatzinformation auskommt.

Die folgenden statistischen Analysen wurden mit S-Plus durchgeführt, vgl. Everitt (1994), Venables & Ripley (1994).

## 3.2 Binäre logistische Regression

Wir reduzieren die Statusinformation über jeden Patienten auf die Alternative

$$St = 0 \text{ (Status = 1)}, \quad St = 1 \text{ (Status = 2 oder 3)}$$

und verzichten zunächst auf die Verwendung der Variablen *Zeit*. Von den 60 Patienten haben 33 überlebt ( $St = 0$ ) und 27 sind verstorben ( $St = 1$ ). Bezeichnen wir die für die Auswertung benutzten Regressoren mit  $z_1, \dots, z_m$  und eine Linearkombination mit

$$\eta = \beta_0 + \beta_1 z_1 + \dots + \beta_m z_m, \quad (11)$$

so lautet die Modellgleichung für den  $i$ -ten Patienten

$$p_i = F(\eta_i), \quad i = 1, \dots, 60. \quad (12)$$

Dabei bezeichnet  $p_i$  die Sterbewahrscheinlichkeit für den  $i$ -ten Patienten;  $\eta_i$  steht für die Linearkombination (11) im Fall des  $i$ -ten Patienten und  $F$  für die logistische Verteilungsfunktion  $F(\eta) = 1/(1 + e^{-\eta})$ .

Nach der Maximum-Likelihood Methode werden die Schätzer  $\hat{\beta}_j$  für die Koeffizienten  $\beta_j$  berechnet sowie ihre zugehörigen (asymptotischen) Standardfehler (standard errors) und Wald-Statistiken  $W = t^2$ . Ferner wird aus den  $\hat{\beta}_0, \dots, \hat{\beta}_m$  gebildet

- (i) log-likelihood  $\ln L$ ,
- (ii) predicted probabilities  $\hat{p}_i$ ,

letzteres für jeden Fall  $i = 1, \dots, 60$ . Führt man (willkürlich) einen Schnittpunkt  $p_c = 0.5$  ein und benennt alle Fälle  $i$  mit  $St_i = 0$  [ $St_i = 1$ ] als "richtig vorhergesagt", falls  $\hat{p}_i \leq p_c$  [ $\hat{p}_i > p_c$ ], so wird mit  $n_0$  [ $n_1$ ] die Anzahl der Fälle in der Gruppe  $St = 0$  [ $St = 1$ ] bezeichnet, die richtig vorhergesagt wurden.

Alle  $m = 8$  Variablen, nämlich die 7 neu entwickelten Maße aus (9) sowie  $\log VESh$ , wurden als mögliche Einflußfaktoren (Regressoren) berücksichtigt. Die besten Ergebnisse, gemessen am Wert der log-likelihood Funktion  $\ln L$ , zeigten bei  $p = 3, 4$  bzw. 6 Regressoren jeweils die folgenden Variablenätze:

$$\begin{aligned}
 p = 3 : & \log VESh, s_z, m_o \\
 & \ln L = -33.79, n_0 = 25 (76\%), n_1 = 18 (67\%) \\
 p = 4 : & \log VESh, s_z, x_u, m_o \\
 & \ln L = -33.39, n_0 = 24 (73\%), n_1 = 19 (70\%) \\
 p = 6 : & \log VESh, s_z, x_o, m_u, m_o, m_z \\
 & \ln L = -32.89, n_0 = 25 (76\%), n_1 = 20 (74\%),
 \end{aligned}$$

vgl. Tab. 3a). Die Sterbewahrscheinlichkeit wird vergrößert durch

kleinere Werte der Variablen  $s_z, m_z, m_o$   
 größere Werte der Variablen  $x_u, m_u$   
 größere Werte der Variablen  $VESh$

Die Analyse mit den 3 Variablen  $\log VESh, \alpha_{VES}$  und  $\alpha_{sin}$  liefert

$$\ln L = -29.72, n_0 = 27 (82\%), n_1 = 18 (67\%)$$

und damit eine etwas bessere Übereinstimmung mit den beobachteten Daten (Tab. 3b)). Histogramme der predicted probabilities  $\hat{p}_i$ , getrennt aufgetragen nach den 33 Fällen mit  $St = 0$  und den 27 Fällen mit  $St = 1$ , bestätigen für beide Ansätze die guten Prädiktionen, vgl. Abb. 5 und 6.

### 3.3 Coxsches Regressionsmodell

Die individuelle Intensitätsfunktion  $\lambda_i(t)$  des (eventuell zensierten) Überlebensprozesses des  $i$ -ten Patienten wird in der Form

$$\lambda_i(t) = \alpha_0(t) \exp(\eta_i) Y_i(t), \quad t \geq 0, \quad (13)$$

eines Cox'schen Regressionsmodells angegeben. Dabei ist  $\alpha_0(t)$  die sog. baseline Hazardfunktion,  $Y_i(t) = 1$ , falls der  $i$ -te Patient zur Zeit  $t$  nach der EKG-Aufnahme noch lebt (bzw. noch unter Beobachtung ist; siehe die Variable *Zeit*), sonst  $Y_i(t) = 0$ , und  $\eta_i$  ist der Wert des Regressionsterms (11) (ohne intercept  $\beta_0$ ). Die partial-likelihood Methode führt zu Schätzern  $\hat{\beta}_j$  sowie den zugehörigen Standardfehlern (standard errors) und Wald-Statistiken  $W = z^2$ . Ferner wird aus den  $\hat{\beta}_j$  berechnet

- (i) die partial-likelihood ratio Statistik  $R = 2(\ln L(\hat{\beta}) - \ln L(\beta^{(0)}))$
- (ii) die Martingal-Residuen  $\hat{m}_i = St_i - \int_0^\infty \lambda_i(s, \hat{\beta}) ds$

letzteres für jeden Fall  $i = 1, \dots, 60$ , mit der dichotomen Statusvariablen  $St_i = 0$  oder 1 wie in 3.2 oben. Aus den Werten  $\hat{m}_i$  wird noch die Residuenquadrat-Summe  $RESSQ = \sum_i \hat{m}_i^2$  abgeleitet.

Wie zuvor wird aus den  $m = 8$  Regressorvariablen der beste Satz mit  $p = 3$ ,  $p = 4$  und  $p = 6$  Variablen ermittelt:

$$\begin{aligned}
 p = 3 : & \log VESh, s_z, m_o \quad (R = 15.2, \text{RESSQ} = 23.8) \\
 p = 4 : & \log VESh, s_z, s_f, m_o \quad (R = 16.9, \text{RESSQ} = 23.4) \\
 p = 6 : & \log VESh, s_z, x_o, m_u, m_o, m_z \quad (R = 20.1, \text{RESSQ} = 23.2),
 \end{aligned}
 \tag{14}$$

vgl. Tab. 4a). Der Ansatz mit den 3 Regressoren  $\log VESh$ ,  $\alpha_{VES}$  und  $\alpha_{sin}$  erzielt die folgenden Werte

$$R = 18.0, \quad \text{RESSQ} = 23.2, \tag{15}$$

vgl. Tab 4b). Alle Ansätze zeigen in etwa die gleiche Anpassungsgüte an die Daten. Bei einem Vergleich untereinander ist jedoch die unterschiedliche Anzahl der verwendeten Parameter zu berücksichtigen. Das Kriterium von Akaike

$$AIC = \ln L - 2 * p$$

weist von den 3 Lösungen (14) diejenige mit  $p = 3$  als die beste aus. Wählt man diese als die Nullhypothese, so ist die Differenz der likelihood-ratio Werte der beiden anderen Lösungen (d. i.  $R = 16.9$  und  $20.1$ ) zum Wert  $R = 15.2$  nicht signifikant: Die Differenzen 1.7 und 4.9 überschreiten nicht die 95%-Quantile 3.84 bzw. 7.82 der  $\chi^2$ -Verteilung mit 1 bzw. 3 Freiheitsgraden. Von den 3 Lösungen (14) ist also diejenige mit  $p = 3$  vorzuziehen.

Damit ergibt das Modell mit den Variablen  $\log VESh$ ,  $\alpha_{VES}$  und  $\alpha_{sin}$  insgesamt die beste Anpassung an die Daten.

Ein graphischer Ansatz, um die Anpassungsgüte des Coxschen Modells mit den verschiedenen Regressorensätzen zu überprüfen, verwendet die integrierten Intensitätsraten

$$\hat{\Lambda}_i = \int_0^{\infty} \lambda_i(s, \hat{\beta}) ds.$$

Die Histogramme der  $\hat{\Lambda}_i$ -Werte wurden dabei getrennt nach dem Status der Patienten aufgetragen (Abb. 7 und 8). Beim Modell mit den  $p = 6$  Variablen aus (14) führt eine Unterteilung beim Wert  $\Lambda = 0.6$  zu einer korrekten Zuordnung von 27 der 33 nicht Verstorbenen und von 12 der 27 Verstorbenen. Beim Modell mit den 3 Merkmalen  $\log VESh$ ,  $\alpha_{VES}$  und  $\alpha_{sin}$  sind (mit einer Unterteilung bei  $\Lambda = 0.5$ ) entsprechend 28 der 33 und 11 der 27 Patienten korrekt zugeordnet.

### 3.4 Zusammenfassung

Zunächst interessiert ein Vergleich des bisherigen Auswertesystems, basierend auf  $\log VESh$ ,  $\alpha_{VES}$  und  $\alpha_{sin}$ , mit dem hier vorgeschlagenen: Während die Variable  $s_z$  die Rolle von  $\alpha_{sin}$  übernimmt ( $\text{cor}(s_z, \alpha_{sin}) = 0.63$ ), gibt es keine Entsprechung zur Variable  $\alpha_{VES}$ . Vielmehr müssen mehrere (mindestens 3) Variablen aus (9) die Funktion von  $\alpha_{VES}$  übernehmen. Zu diesem Nachteil gesellen sich aber mindestens drei Vorteile des hier vorgeschlagenen Auswertesystems:

1. Die Prädiktoren (9) können vollständig automatisch (algorithmisch) aus den Ausgangsdaten (1) gewonnen werden.

2. Bei der Bestimmung der Prädiktoren reichen die aufeinanderfolgenden Zeitabstände (1) aus; es bedarf keiner zusätzlichen Klassifikation der Herzschläge.
3. Die Analyse mit der hier beschriebenen Methode ist nach Ablauf eines Teils der Aufnahmezeit – etwa nach 12 Stunden – im Prinzip online möglich. Die beiden Parameter  $\alpha_{sin}$  und  $\alpha_{VES}$  dagegen können erst nach Vorliegen aller RR-Intervalle, d. h. nach 24 Std., bestimmt werden.

Über Punkt 2. hinaus scheint das vorgeschlagene System sogar auf die Variable  $VESh$  verzichten zu können. Dazu führen wir – als Ersatz für  $\log VESh$  – die Variable

$$\log \frac{N_f}{N},$$

ein, wobei  $N_f$  wie oben die Anzahl der  $t_i$  bezeichnet, die in einen der beiden Flügel des Histogramms fallen (es ist  $\text{cor}(\log(VESh), \log(N_f/N)) = 0.75$ ). Dann erhalten wir für den – um  $\log(N_f/N)$  vermehrten – Variablensatz (9)

$$\begin{aligned} \text{logistische Regression: } \ln L &= -33.25, \quad n_0 = 24, \quad n_1 = 17 \\ \text{Coxsche Regression: } R &= 19.2, \quad RESSQ = 24.19. \end{aligned}$$

Zumindest für das Coxsche Modell entspricht dieses Resultat etwa den Ergebnissen, die unter Benutzung der Variablen  $VESh$  zustande kamen, nämlich (14) und (15).

Die erwähnten Vorteile lassen es angebracht erscheinen, den hier vorgeschlagene Ansatz in Zukunft weiter zu entwickeln.

## Anerkennung

Diese Arbeit wurde zum Teil von der Deutschen Forschungsgemeinschaft (SFB 386) unterstützt.

## Literatur

CAST Investigators: The Cardiac Arrhythmia Suppression Trial (CAST). N Engl J Med 321 (1989), 406 - 412.

Everitt, B. S.: A Handbook of Statistical Analysis using S-Plus. Chapman & Hall, London (1994).

Schmidt, G., Morfill, G. E. et al.: Variability of Ventricular Premature Complexes and Mortality Risk. PACE 19 (1996), 976 - 980.

Silverman, B. W.: Density Estimation for Statistical and Data Analysis. Chapman & Hall, London (1986).

Moss, A. J., Davis, H. T. et al.: Ventricular ectopic beats and their relation to sudden and non sudden cardiac death after myocard infarction. Circulation 60 (1979), 998 - 1003.

Trappe, H.-J., Heintze, J., Lichtlen P. R.: Identifikation des rhythmusgefährdeten Patienten. Internist 37 (1996), 34 - 44.

Venables, W. N., Ripley, B. D.: Modern Applied Statistics with S-Plus. Springer, New York (1994).

## Anschriften

Helmut Pruscha. Mathematisches Institut der Universität München, Theresienstr. 39, D-80333 München, pruscha@rz.mathematik.uni-muenchen.de

Kurt Ulm. Institut für Medizinische Statistik der TU München, Ismaninger Str. 22, D-81675 München, ulm@imse.med.tu-muenchen.de

Georg Schmidt. I. Medizinische Klinik der TU München, Klinikum rechts der Isar

TABELLE 1

	Sta	N	Nf	Sf*Sf	Sz	Xu	Mu	Xo	Mo	Mz
OTDJFZ	1	126978	976	0.016	0.0212	0.875	0.078	1.125	0.075	0.77
OYEPCF	1	87227	59	0.351	0.1573	0.475	0.001	1.775	0.054	13.93
OTEOGR	1	70633	3019	0.334	0.1071	0.675	1.126	1.325	1.014	2.83
OTSYBS	1	64452	15842	0.217	0.1423	0.775	4.583	1.175	4.037	2.35
OTEMQL	1	68749	2587	0.284	0.1162	0.775	1.074	1.375	0.583	1.66
OTDLWV	3	65461	2033	0.282	0.0886	0.775	0.754	1.225	0.533	1.35
OUCVUR	1	87474	5324	0.080	0.0916	0.825	1.098	1.225	1.053	1.47
OWABPG	1	89960	42310	0.173	0.1115	0.825	9.740	1.175	10.285	1.56
OTAMWD	1	57718	2912	0.219	0.1805	0.775	1.048	1.225	1.051	2.28
OWKWXA	1	105695	4279	0.049	0.0588	0.825	0.678	1.175	0.547	1.13
PEVOZZ	1	104625	8502	0.073	0.0944	0.825	1.356	1.175	1.196	1.44
OTWMRD	1	82178	50015	0.162	0.2018	0.875	10.991	1.175	9.467	1.82
OTHPSO	2	83564	18941	0.086	0.0787	0.825	3.526	1.175	3.726	1.53
PBROBW	1	98129	16605	0.109	0.1102	0.825	3.188	1.225	2.625	2.41
OVXSBH	1	61389	1331	0.216	0.1964	0.825	0.413	1.175	0.293	1.52
OVZYLC	1	72272	6564	0.293	0.2122	0.725	2.672	1.325	1.310	4.10
OTOXPP	2	110771	17620	0.013	0.0586	0.825	2.412	1.175	1.774	2.01
OTRKKO	1	74760	15848	0.107	0.1205	0.825	3.848	1.225	3.614	2.14
PENHVE	1	97684	5029	0.158	0.1281	0.725	1.263	1.325	0.975	2.21
OVWURL	1	80007	349	0.272	0.1089	0.575	0.167	1.425	0.060	3.95
OTTTMZ	2	110656	1455	0.052	0.0803	0.825	0.192	1.175	0.154	1.16
OUVDPV	1	97395	12544	0.105	0.0448	0.775	2.550	1.225	2.505	1.96
PEJFOM	3	86534	9124	0.207	0.1592	0.775	2.265	1.225	2.079	1.84
PAQKKU	1	68374	4445	0.205	0.1521	0.825	1.238	1.225	0.977	1.99
PFBGBE	3	120653	9477	0.049	0.0676	0.875	2.048	1.325	0.151	2.81
OTOBPY	2	141438	3931	0.003	0.0300	0.825	0.504	1.225	0.399	0.93
OUPGAY	1	100707	653	0.011	0.0278	0.875	0.078	1.125	0.074	0.67
PFQWGU	3	93777	3013	0.188	0.1315	0.775	0.796	1.325	0.669	1.49
OWEGSQ	1	102197	954	0.090	0.1242	0.825	0.178	1.175	0.142	0.71
PBMTAO	1	100556	33994	0.069	0.1068	0.825	5.530	1.175	5.874	1.81
OWCPKA	1	87837	6389	0.031	0.3149	0.775	2.221	1.675	0.008	6.06
PFLYRH	2	109348	2747	0.077	0.0841	0.825	0.461	1.225	0.396	1.13
PFLZLP	1	112605	3830	0.003	0.0930	0.775	0.667	1.225	0.640	1.63
OTTTZH	1	71721	878	0.139	0.1451	0.775	0.258	1.225	0.228	2.22
PFSMAS	1	99143	44607	0.116	0.1638	0.875	7.475	1.175	6.403	1.94

PEUUUR	1	81087	2740	0.210	0.1443	0.725	0.725	1.225	0.687	1.51
PFVUUO	1	81912	5003	0.218	0.1908	0.775	1.633	1.375	0.924	4.14
OTHOTU	2	149995	2970	0.028	0.1023	0.775	0.614	1.475	0.016	3.08
OWATHB	2	100448	11311	0.081	0.0966	0.825	1.960	1.225	1.610	2.30
PBNLNM	3	108318	4514	0.052	0.0612	0.825	0.635	1.175	0.587	1.19
PEHPSF	3	71314	1316	0.216	0.1295	0.825	0.337	1.175	0.228	1.78
OTTXZX	1	69353	5003	0.092	0.1281	0.825	1.004	1.175	1.074	1.47
OZNNWS	3	87853	3956	0.171	0.0952	0.775	0.989	1.275	0.818	1.84
PEURRU	3	115912	9695	0.063	0.1046	0.775	2.786	1.825	0.005	6.38
PAJCPO	3	104711	26307	0.082	0.1204	0.875	3.752	1.175	3.567	1.39
PFRSVW	2	97984	17010	0.145	0.1812	0.775	6.483	1.975	0.017	9.34
PFXKKI	3	95974	18081	0.099	0.1082	0.825	3.277	1.175	3.088	2.02
OZXPMP	2	102698	2661	0.162	0.1325	0.725	0.418	1.275	0.598	2.00
PFVVYC	3	96594	12528	0.076	0.0993	0.825	1.867	1.175	1.879	1.81
OUVCVT	1	109348	2278	0.203	0.0807	0.825	0.778	1.325	0.173	1.58
OVZXPX	1	103107	4187	0.070	0.1164	0.825	0.550	1.175	0.528	2.04
OZYLXN	3	104987	11101	0.092	0.0985	0.825	1.952	1.225	1.499	2.21
PEODFG	3	93821	15351	0.116	0.1249	0.825	3.261	1.225	2.818	2.23
PFFKOJ	1	79885	1165	0.092	0.0991	0.825	0.231	1.175	0.223	1.07
PECQXO	3	103495	5705	0.109	0.1739	0.825	1.038	1.225	0.853	1.77
PBPZOP	3	102964	5229	0.083	0.0666	0.825	0.830	1.175	0.787	1.30
PGDDSX	2	114482	17849	0.069	0.0932	0.825	2.451	1.175	2.440	1.36
PBQSPP	1	90263	282	0.070	0.1122	0.775	0.050	1.175	0.042	1.79
PEZQBO	2	85605	6026	0.094	0.1269	0.825	1.206	1.225	0.947	1.97
PEVQEJ	2	79067	403	0.102	0.1130	0.775	0.088	1.175	0.062	1.32

Tab. 1. 60 Patienten mit Oxfordkürzel, Status und den in 3.1 erläuterten Variablen.

TABELLE 2

	Sta	Zeit	VESh	aSin	aVES
OTDJFZ	1	1.36	29.43	0.01	0.09
OYEPCF	1	0.45	13.00	0.80	0.84
OTEOGR	1	0.91	63.55	0.24	0.42
OTSYBS	1	0.67	27.39	0.31	0.87
OTEMQL	1	0.52	30.33	0.17	0.93
OTDLWV	3	0.19	10.43	0.10	1.50
OUCVUR	1	1.88	14.65	0.17	1.47
OWABPG	1	9.76	659.00	0.44	0.42
OTAMWD	1	0.84	53.30	0.46	1.32
OWKWXA	1	0.26	33.16	0.08	1.56
PEVOZZ	1	11.25	70.70	0.18	1.38
OTWMRD	1	1.75	178.96	0.80	1.08
OTHPSO	2	5.30	407.26	0.29	0.81
PBROBW	1	5.65	97.90	0.35	1.32
OVXSBH	1	0.27	16.55	0.33	1.92
OVZYLC	1	7.70	185.00	0.34	1.14
OTOXPP	2	9.13	7.35	0.15	2.22
OTRKKO	1	0.42	330.52	0.45	0.96
PENVHE	1	8.90	66.86	0.18	1.50
OVWURL	1	0.23	13.48	0.17	2.07
OTTTMZ	2	1.54	18.74	0.11	1.98
OUVDPV	1	0.18	291.05	0.12	1.08
PEJFOM	3	0.25	119.40	0.29	1.38
PAQKKU	1	7.18	84.70	0.33	1.50
PFBGBE	3	0.29	300.00	0.13	1.08
OTOBPY	2	0.01	65.70	0.01	1.59
OUPGAY	1	0.10	10.05	0.01	2.22
PFQWGU	3	2.81	68.70	0.30	1.62
OWEGSQ	1	0.29	15.71	0.22	2.13
PBMTAO	1	5.56	809.90	0.50	0.84
OWCPKA	1	0.23	87.43	0.27	1.59
PFLYRH	2	0.68	55.80	0.11	1.74
PFLZLP	1	8.04	83.43	0.17	1.62
OTTTZH	1	1.78	22.11	0.27	2.07
PFSMAS	1	8.08	898.90	0.66	0.84

PEUUUR	1	7.01	42.60	0.35	1.86
PFVUUO	1	7.10	150.30	0.46	1.44
OTHOTU	2	0.08	74.48	0.09	1.71
OWATHB	2	6.57	245.00	0.23	1.32
PBNLNM	3	1.11	99.80	0.08	1.62
PEHPSF	3	0.26	6.60	0.25	2.55
OTTXZX	1	1.78	157.53	0.35	1.50
OZNNWS	3	0.43	112.30	0.16	1.62
PEURRU	3	1.45	692.50	0.18	1.02
PAJCPO	3	1.71	484.70	0.32	1.14
PFRSVW	2	0.92	351.60	0.51	1.26
PFXKKI	3	6.07	431.70	0.33	1.20
OZXPMP	2	2.50	43.70	0.28	1.98
PFVVYC	3	0.57	322.36	0.22	1.32
OUVCVT	1	0.18	29.57	0.12	2.13
OVZXPX	1	6.28	99.00	0.26	1.74
OZYLXN	3	1.25	254.10	0.25	1.44
PEODFG	3	0.34	319.60	0.38	1.38
PFFKOJ	1	8.90	26.30	0.18	2.22
PECQXO	3	4.35	163.90	0.21	1.62
PBPZOP	3	4.90	99.50	0.10	1.80
PGDDSX	2	1.92	353.80	0.26	1.38
PBQSPP	1	8.72	4.81	0.22	2.82
PEZQBO	2	4.28	343.80	0.32	1.44
PEVQEJ	2	2.05	21.20	0.20	2.58

Tab. 2. Dieselben 60 Patienten mit Oxfordkürzel, Status,  $\alpha_{sin}$ ,  $\alpha_{VES}$  und den in (10) erläuterten Variablen Zeit und VESh.

TABELLE 3

a)

	Value	Std. Error	t value
(Intercept)	-1.9132830	1.3426935	-1.424959
logV	0.9207082	0.3300972	2.789203
Mo	-0.5749548	0.2751144	-2.089876
Sz	-13.4617849	6.7868838	-1.983500

Residual Deviance: 67.58086 on 56 degrees of freedom  
 \$loglike: -33.79043 \$correctly classified \$n0: 25 \$n1: 18

	Value	Std. Error	t value
(Intercept)	-6.3657088	5.6093183	-1.1348454
logV	0.8716909	0.3317522	2.6275361
Mo	-0.6059389	0.2784662	-2.1759876
Sz	-11.0558965	7.0504525	-1.5681116
Xu	5.5481580	6.6898576	0.8293387

Residual Deviance: 66.78382 on 55 degrees of freedom  
 \$loglike: -33.39191 \$n0: 24 \$n1: 19

	Value	Std. Error	t value
(Intercept)	-2.5150422	6.8282655	-0.3683281
logV	0.9093033	0.3679296	2.4714058
Mo	-1.4317593	0.9615199	-1.4890584
Sz	-10.5738736	8.3680475	-1.2636010
Mu	0.8862979	0.8695385	1.0192739
Mz	-0.6435382	0.7311806	-0.8801359
Xo	1.0464480	5.9623647	0.1755089

Residual Deviance: 65.79095 on 53 degrees of freedom  
 \$loglike: -32.89548 \$n0: 25 \$n1: 20

-----



Likelihood ratio test= 16.9 on 4 df, p=0.00204  
 Efficient score test = 12.8 on 4 df, p=0.0124 ressq 23.4255

	coef	exp(coef)	se(coef)	z	p
Mu	0.943	2.57e+00	0.535	1.76	0.0782
Mo	-1.258	2.84e-01	0.572	-2.20	0.0278
Sz	-11.241	1.31e-05	6.488	-1.73	0.0832
Xo	5.929	3.76e+02	3.861	1.54	0.1246
Mz	-1.215	2.97e-01	0.620	-1.96	0.0501
logV	0.486	1.63e+00	0.245	1.98	0.0477

Likelihood ratio test= 20.1 on 6 df, p=0.00268  
 Efficient score test = 18.3 on 6 df, p=0.00562 ressq 23.2471

-----  
 b)

	coef	exp(coef)	se(coef)	z	p
aVES	1.353	3.868778	0.848	1.59	0.11079
aSin	-7.346	0.000645	2.270	-3.24	0.00121
logV	0.848	2.334871	0.328	2.58	0.00975

Likelihood ratio test= 18 on 3 df, p=0.000436  
 Efficient score test = 13.2 on 3 df, p=0.00419 ressq 23.2634  
 -----

Tab. 4. Ergebnisse der Coxschen Regressionsanalyse, für die Daten der Tab. 1 (a, oben) einschließlich der Variablen VESh, und für die Daten der Tab. 2 (b, unten).  $\log V = \log(\text{VESh})$ .

Histogramm der Zeitintervalle

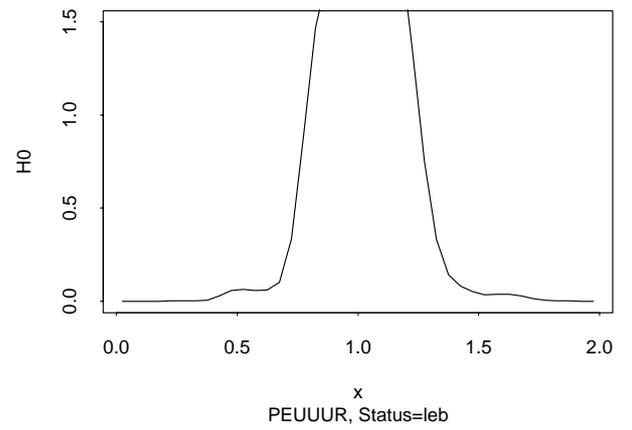
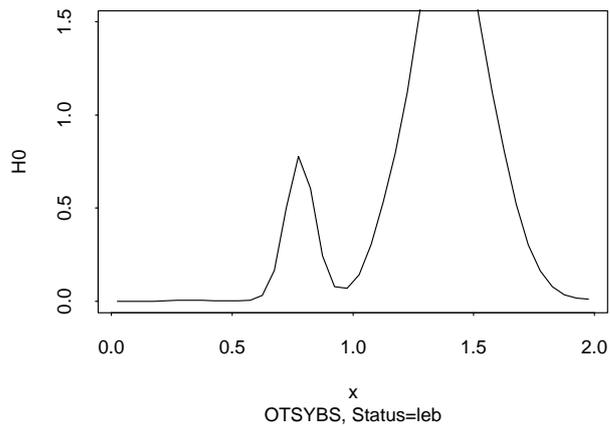
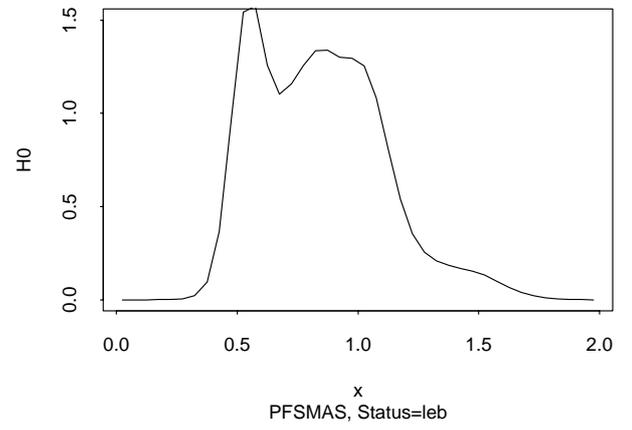
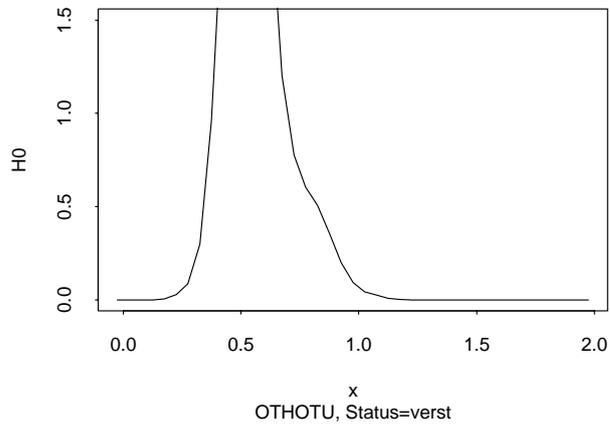


Abbildung 1: Geglättete Histogramme der Originaldaten  $x_i$  für 4 ausgewählte Patienten.

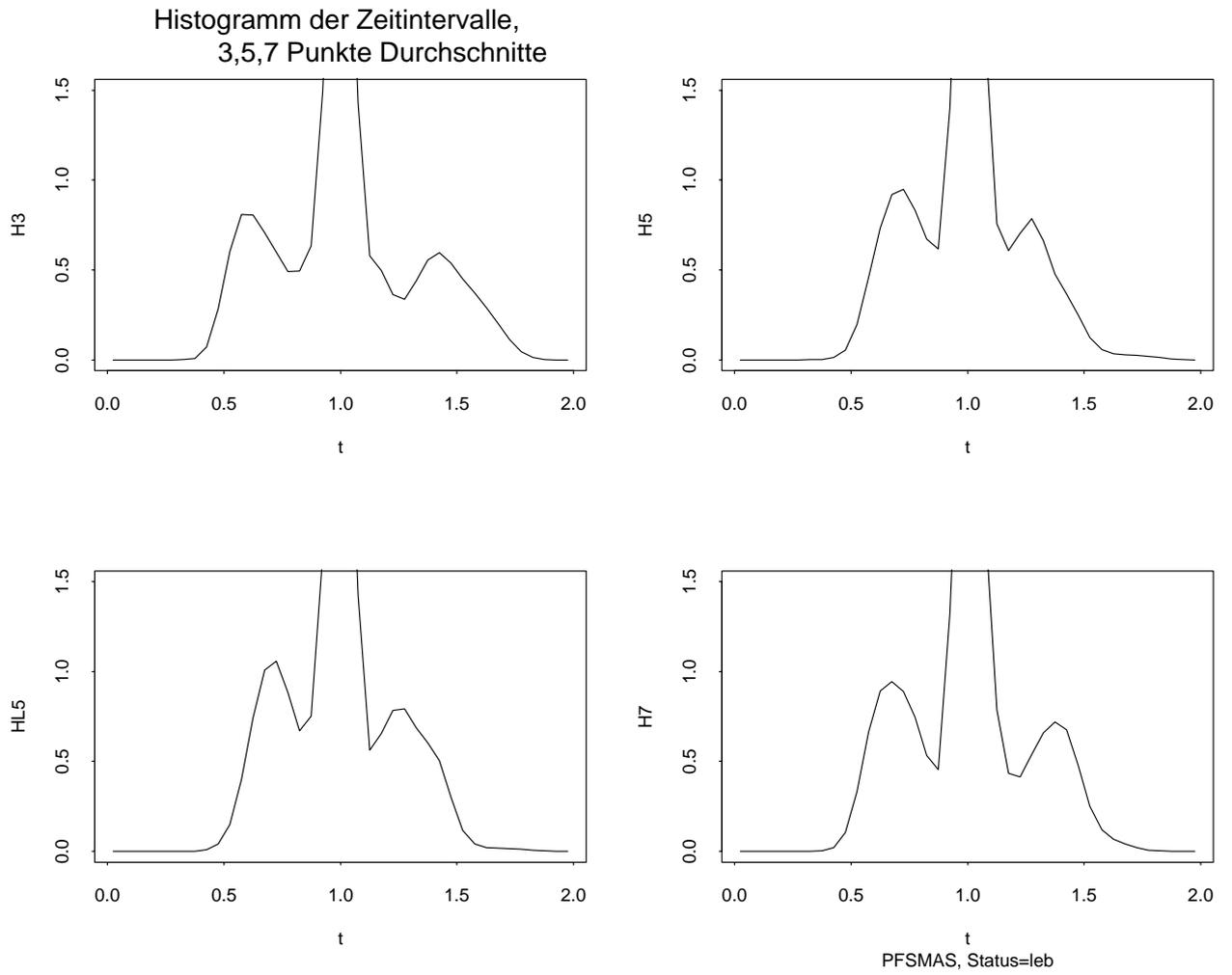


Abbildung 2: Geglättete Histogramme der transformierten Daten  $t_i$  für den Fall PFSMAS, jeweils mit  $m = 3, 5$  und  $7$  Punkte-Durchschnitt (H3, H5, H7), und einmal mit  $m = 5$  und linksseitigen Durchschnitten (HL5).

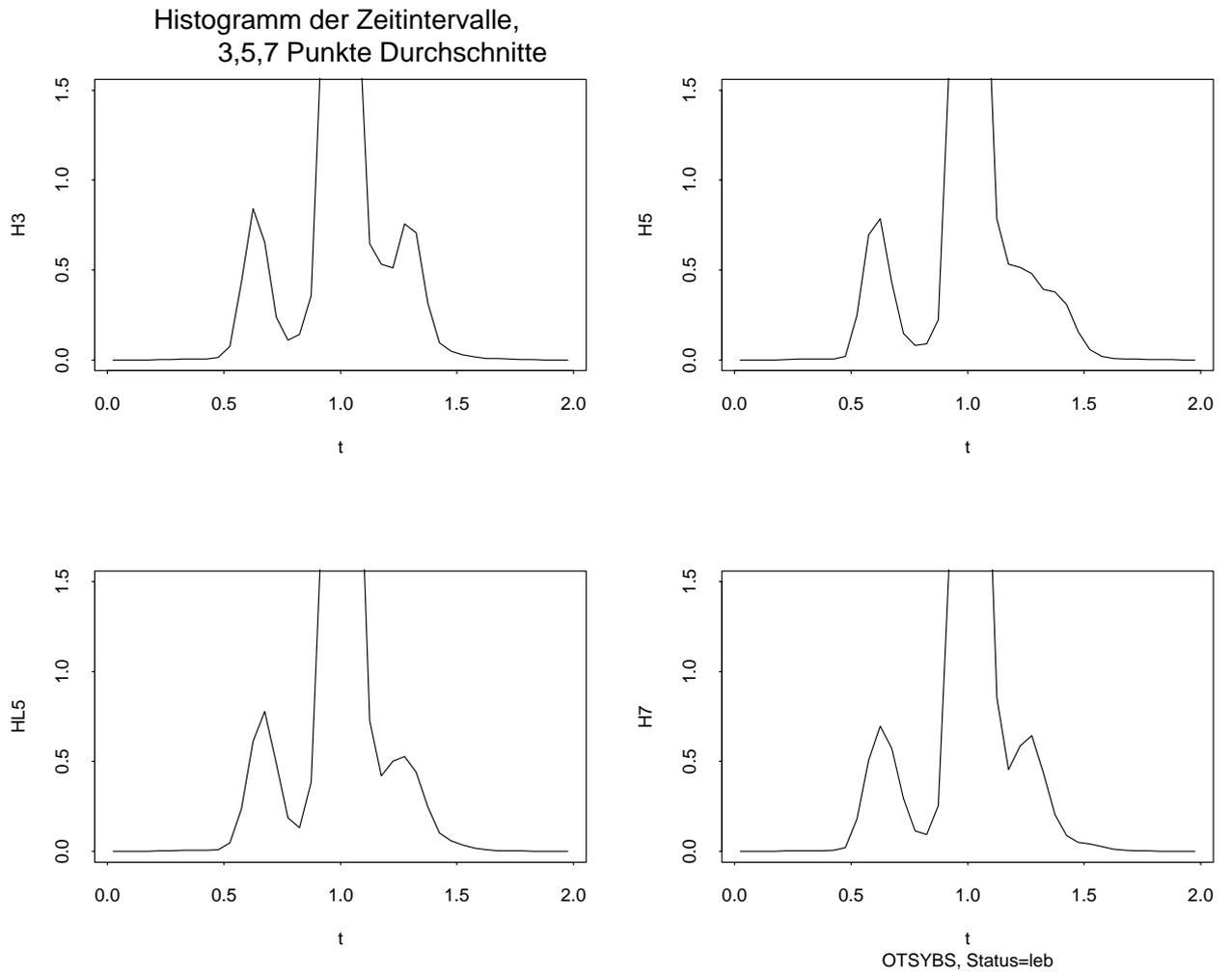


Abbildung 3: Geglättete Histogramme der transformierten Daten  $t_i$  für den Fall OTSYBS, jeweils mit  $m = 3, 5$  und  $7$  Punkte-Durchschnitt (H3, H5, H7), und einmal mit  $m = 5$  und linksseitigen Durchschnitten (HL5).

### Bestimmung der Cutpoints $x_u$ und $x_o$

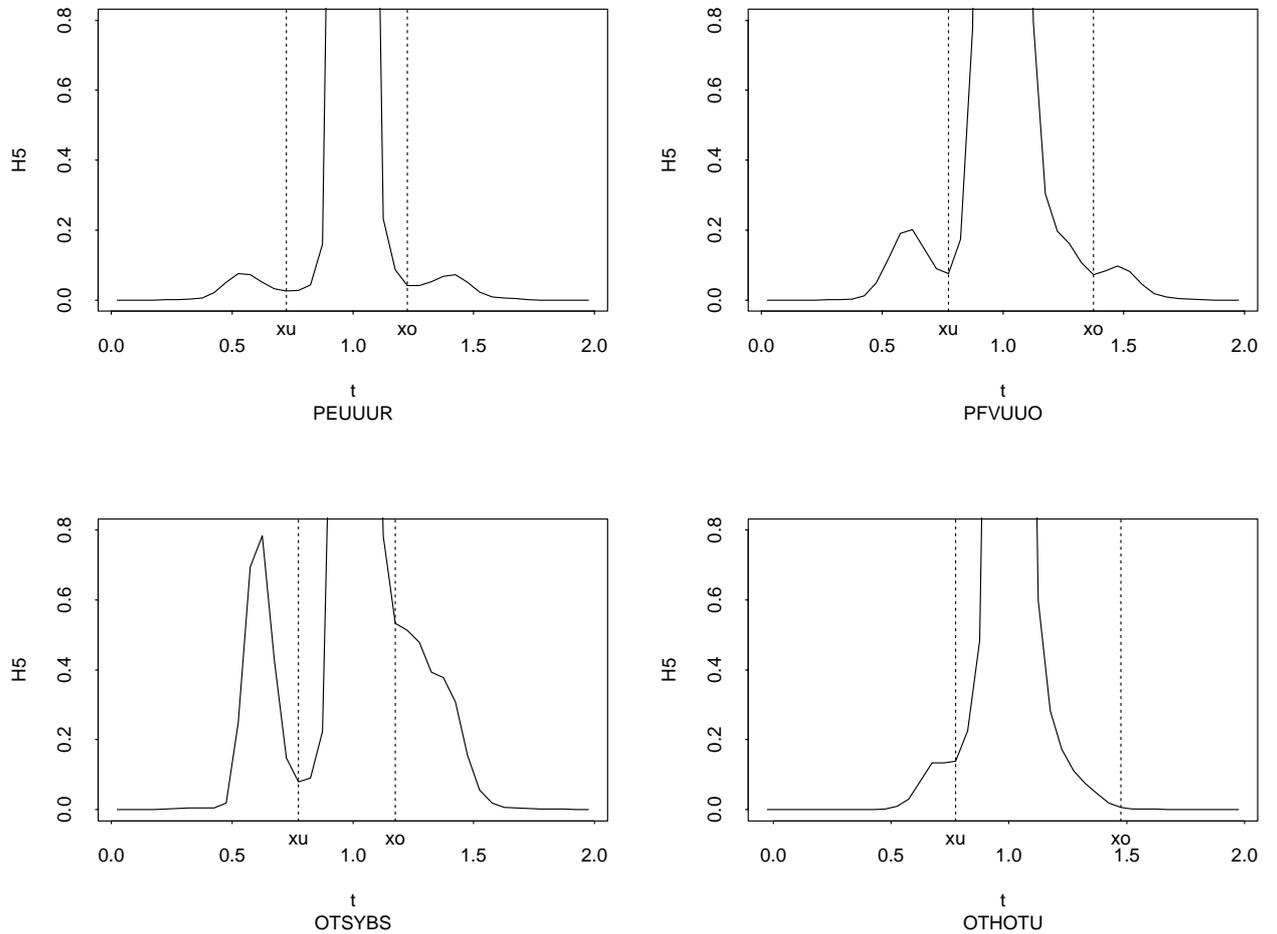


Abbildung 4: Bestimmung der Cutpoint-Werte  $x_u$ ,  $x_o$  aus dem geglätteten Histogramm  $H(t)$ ,  $0 \leq t \leq 2$ . a) Eindeutige Minima unterhalb und oberhalb von  $t = 1$  (Fälle PEUUUR, PFVUUO), b) Krümmungspunkte unterhalb oder oberhalb von  $t = 1$  (Fälle OTSYBS, OTHOTU).

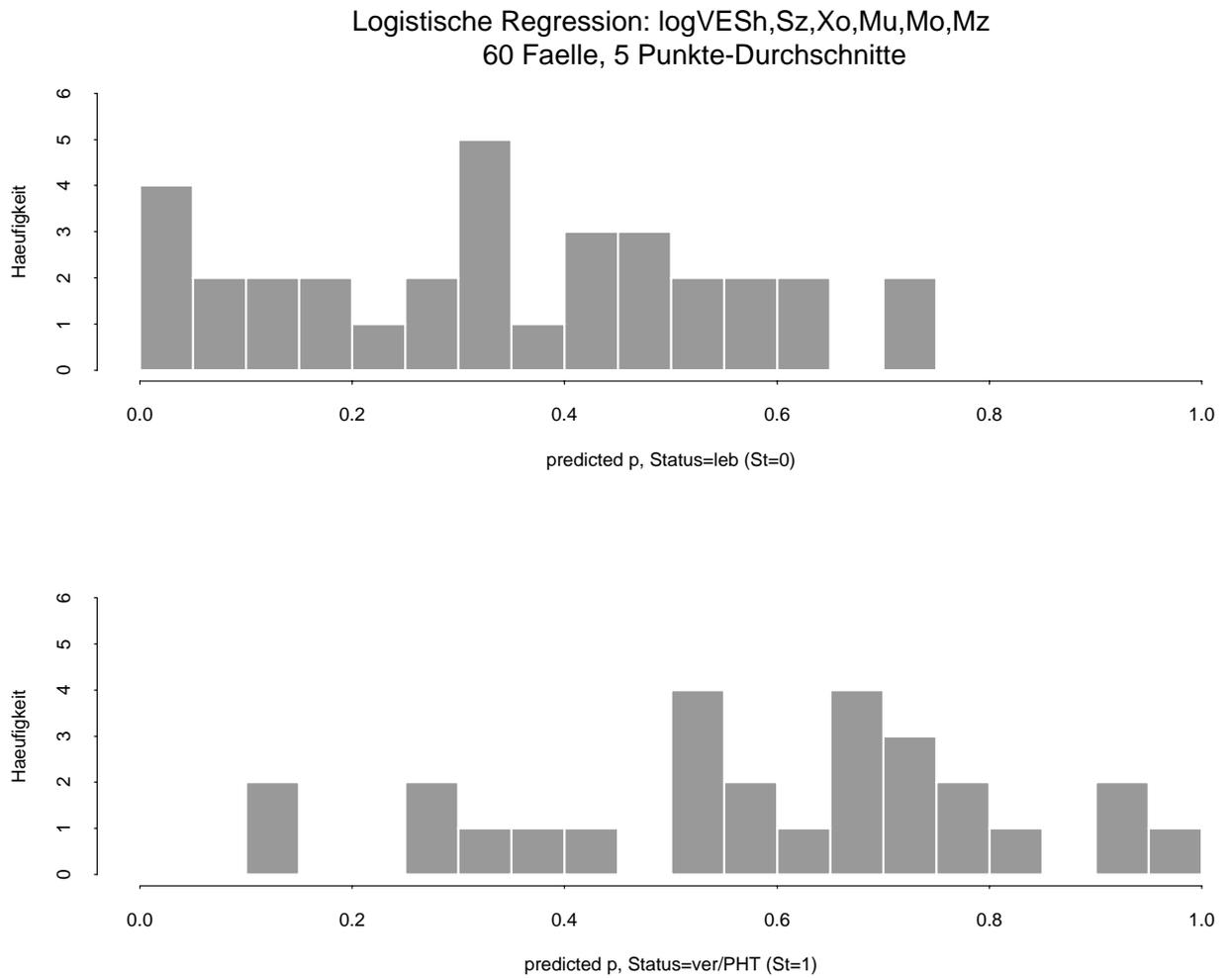


Abbildung 5: Histogramme der (predicted) Wahrscheinlichkeiten  $\hat{p}_i$ , getrennt aufgetragen nach den zwei Gruppen  $St = 0$  (33 Patienten) und  $St = 1$  (27 Patienten). Daten der Tab. 1, mit den Regressoren  $\log VESh$ ,  $s_z$ ,  $x_0$ ,  $m_u$ ,  $m_o$  und  $m_z$ .

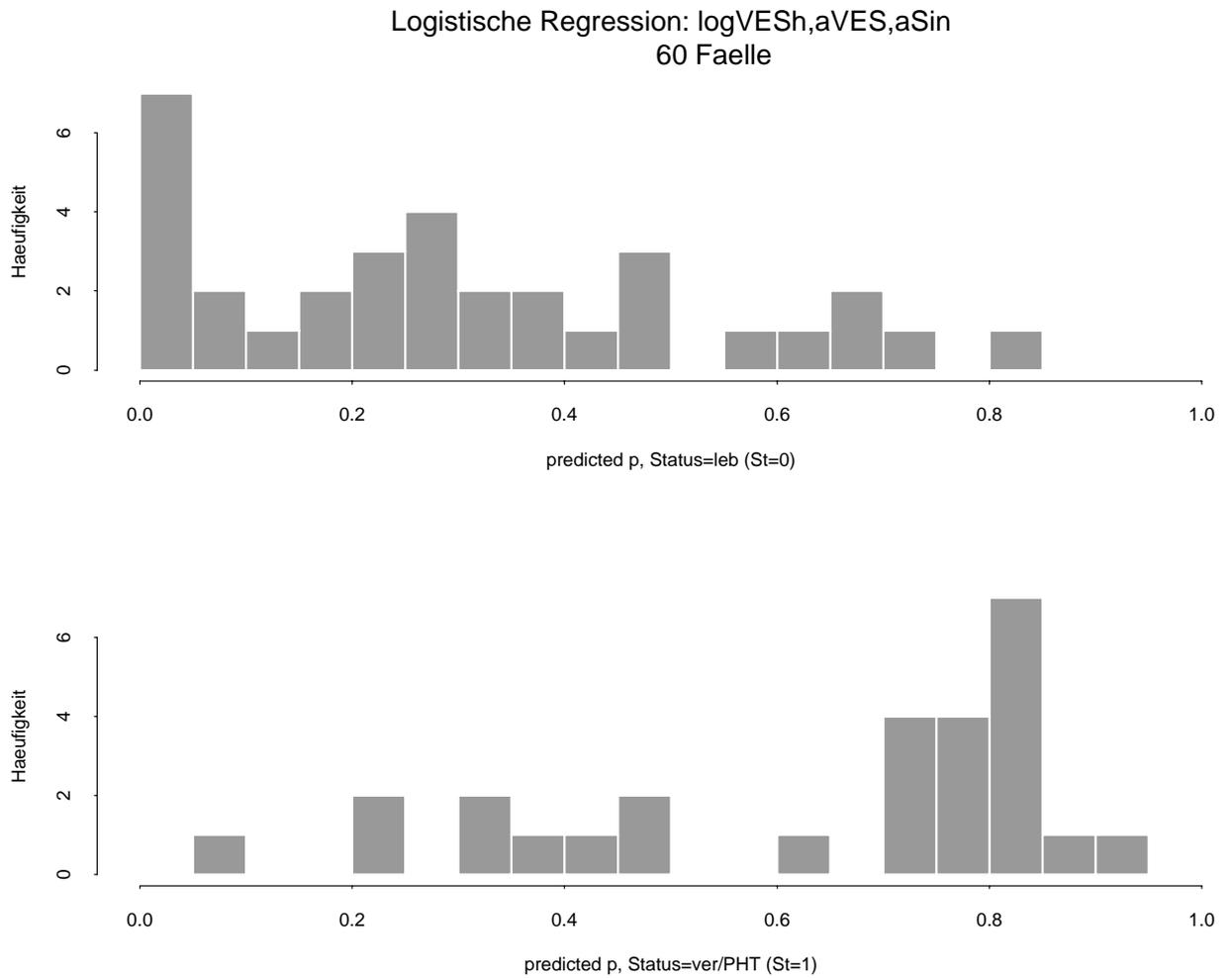


Abbildung 6: Histogramme der (predicted) Wahrscheinlichkeiten  $\hat{p}_i$ , getrennt aufgetragen nach den zwei Gruppen  $St = 0$  (33 Patienten) und  $St = 1$  (27 Patienten). Daten der Tab. 2, mit den Regressoren  $\log VESh$ ,  $\alpha_{VES}$  und  $\alpha_{sin}$ .

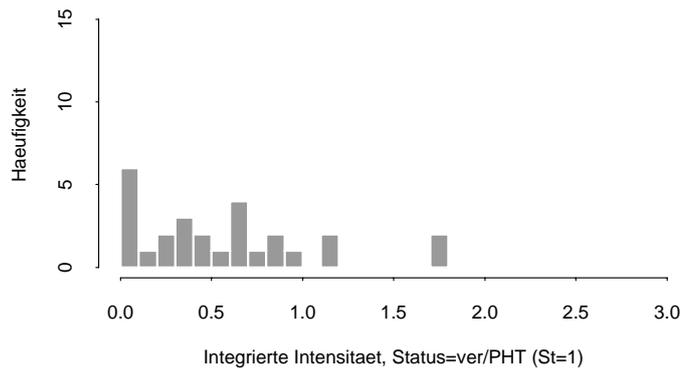
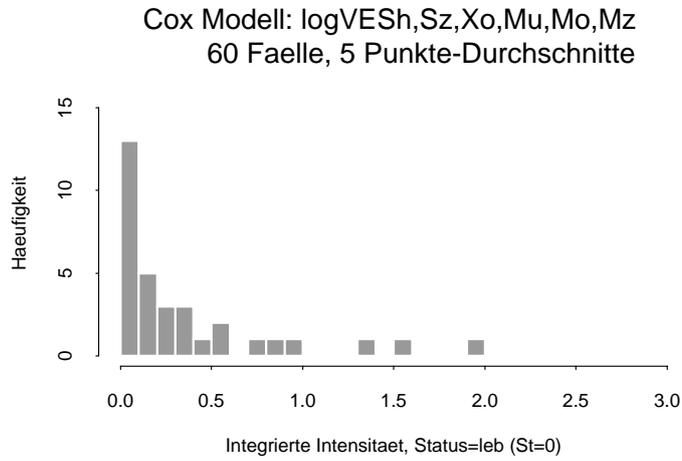


Abbildung 7: Histogramme der integrierten Intensität  $\hat{\Lambda}_i$ , getrennt aufgetragen nach den zwei Gruppen  $St = 0$  (33 Patienten) und  $St = 1$  (27 Patienten). Daten der Tab. 1, mit den Regressoren  $\log VESh$ ,  $s_z$ ,  $x_0$ ,  $m_u$ ,  $m_o$  und  $m_z$ .

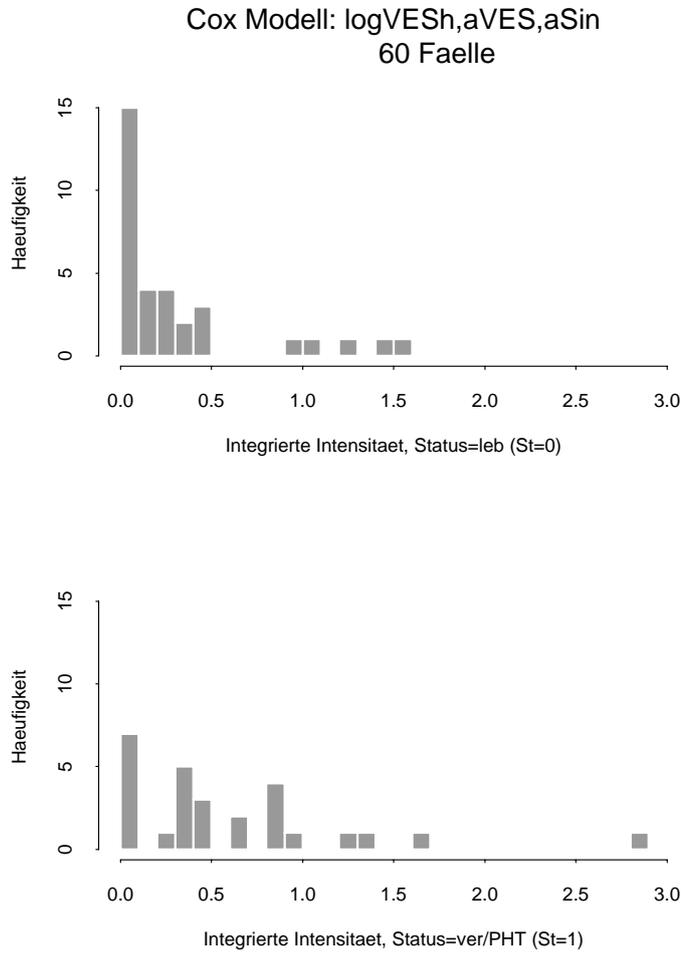


Abbildung 8: Histogramme der integrierten Intensitaet  $\hat{\Lambda}_i$ , getrennt aufgetragen nach den zwei Gruppen  $St = 0$  (33 Patienten) und  $St = 1$  (27 Patienten). Daten der Tab. 2, mit den Regressoren  $\log VESh$ ,  $\alpha_{VES}$  und  $\alpha_{sin}$ .