



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) DE 10 2005 004 294 A1 2006.08.03

(12)

Offenlegungsschrift

(21) Aktenzeichen: 10 2005 004 294.5

(22) Anmeldetag: 28.01.2005

(43) Offenlegungstag: 03.08.2006

(51) Int Cl.⁸: G06F 17/18 (2006.01)

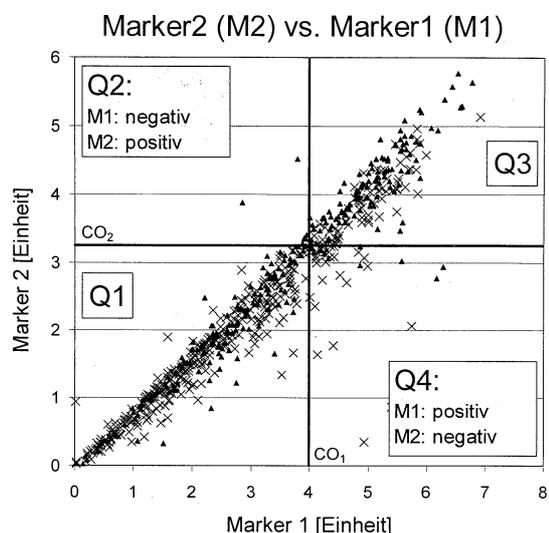
(71) Anmelder:
Bayer HealthCare AG, 51373 Leverkusen, DE

(72) Erfinder:
Butz, Hermann, Dr., 35423 Lich, DE; Keller,
Thomas, Dr., 04275 Leipzig, DE

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) Bezeichnung: **Auswahl und Bewertung diagnostischer Tests mittels Diskordanzanalysen-Charakteristik (DAC)**

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Ermittlung der diagnostischen Güte von diagnostischen Tests anhand von Messwerten, von zwei diagnostischen Tests gemessen an Patienten mit der (positiv) und ohne die (negativ) zu diagnostizierende Erkrankung. Es werden Cut-Off-Paare $(CO_1, CO_2)_k$ nach dem Kriterium gleicher Sensitivität für die Messwerte beider Tests festgelegt und für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ die Messerte ausgewählt, die entweder für den ersten Test oberhalb des zugehörigen Cut-Off (CO_1) und für den zweiten Test unterhalb des zugehörigen Cut-Off (CO_2) liegen, oder die für den ersten Test unterhalb des zugehörigen Cut-Off (CO_1) und für den zweiten Test oberhalb des zugehörigen Cut-Off (CO_2) liegen. Für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ und für die ausgewählten Messwerte wird die Anzahl von denjenigen Messwerten bestimmt, die zu Patienten gehören, die richtig testpositiv, falsch testpositiv, richtig testnegativ und falsch testnegativ getestet wurden. Aus diesen Anzahlen werden die Spezifitäten für beide diagnostischen Tests berechnet.



Beschreibung

[0001] Die Erfindung betrifft die Anwendung eines mathematischen Verfahrens, mit dessen Hilfe ein Verfahren zur Diagnose einer Erkrankung ausgewählt und bewertet werden kann. Die Auswahl und Bewertung eines diagnostischen Verfahrens erfolgt auf der Basis verschiedener Kenngrößen. Im Vordergrund stehen dabei

- die Laboreigenschaften (Richtigkeit, Präzision, Selektivität etc. bei der Erzeugung des Messwertes),
- die diagnostische Güte (Sensitivität, Spezifität der Entscheidung positiv/negativ, berechnet anhand der Anzahlen für richtig positive, richtig negative, falsch positive und falsch negative Entscheidungen.),
- der Kostenfaktor (monetäre Kosten, „Kosten“, d.h. Belastung (z.B. invasive Folge- und Zusatzuntersuchungen für den Patienten) sowie
- die Wirkung (verbesserte Therapiemöglichkeiten, verbessertes Überleben aufgrund der Diagnostik).

[0002] Die Erfindung betrifft die Auswahl bezüglich der diagnostischen Güte. Darüber hinaus sind auch Aussagen zu den „Kosten“ möglich, wobei ein direkter Link zwischen Diagnostik-Bewertung und Kostenbewertung hergestellt wird. Nach dem Stand der Technik werden auf der Basis der Sensitivitäts- und Spezifitätswerte für verschiedene Cut-Offs (Entscheidungsschwellen) des oder der zu untersuchenden Tests zwei Methoden angewandt:

Stand der Technik

- a. Punktweise Analyse von Sensitivität und Spezifität, Vergleich mittels McNemar-Tests:

[0003] Für vorgegebene Cut-Off-Werte werden Sensitivität und Spezifität ermittelt und angegeben. Alternativ können für vorgegebene Werte der Spezifität (bzw. Sensitivität), z.B. 95%, die zugehörigen Werte für die Sensitivität (bzw. Spezifität) ermittelt werden. Werden Sensitivität bzw. Spezifität auf letztere Weise ermittelt, so können deren Werte für zwei diagnostische Tests durch Eintrag der jeweiligen Testergebnisse positiv/negativ in eine Vierfeldertafel verglichen werden. Die Felder beinhalten die Kombinationen „beide Tests positiv“, „beide Tests negativ“ sowie die beiden diskordanten Fälle. Die statistische Auswertung hinsichtlich von Unterschieden von Sensitivität bzw. Spezifität erfolgt mit dem McNemar-Test (Altman DG. Practical statistics for medical research, London: Chapman & Hall, 1991:p416).

- b. ROC (Receiver Operating Characteristics)-Analyse:

[0004] Zur ROC-Analyse werden die Sensitivi-

tät/Spezifitäts-Wertepaare für alle denkbaren Cut-Offs in der Form Richtig-Positiv-Rate = $f(\text{Falsch-Positiv-Rate})$ oder anders formuliert Sensitivität = $f(1 - \text{Spezifität})$ aufgetragen. Das heißt, alle Werte des Wertebereiches (Messbereiches) werden als Cut-Off-Wert verwendet, und die zugehörigen Sensitivitäten und Spezifitäten ermittelt. Als Kenngröße zum Methodenvergleich wird das Integral unter der Funktion, die AUC (Area Under Curve), verwendet. Eine höhere AUC geht mit einer verbesserten diagnostischen Güte einher. Die AUC kann Werte zwischen 0,5 (fehlende Trennschärfe des Tests) und 1 (vollständige Trennschärfe des Tests) annehmen. Der statistischen Bewertung einer AUC mittels z.B. des Wilcoxon-Tests oder der Vergleich mehrerer AUC und die Berechnung einer statistischen Signifikanz für den Unterschied durch die z.B. in Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36 oder in DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45 angegebenen Verfahren sind als mathematische Methoden etabliert.

[0005] Die Ergebnisse der punktwisen Analysen hängen stark von den spezifisch zugrundeliegenden Daten und Verteilungen ab. Damit wird ein Vergleich zwischen verschiedenen Datenerhebungen zu ein und demselben diagnostischen Tests und ein Vergleich zwischen verschiedenen diagnostischen Verfahren schwierig. Punktweise Analysen werden insbesondere problematisch, wenn innerhalb der Datenerhebungen nur Teilbereiche des durch die Grundgesamtheit beschriebenen Wertebereiches untersucht werden. Bei teuren Messverfahren ist es beispielsweise üblich, nur Patienten in eine Datenerhebung einzuschließen, die Messwerte um die Entscheidungsgrenze, den Cut-Off-Wert, aufweisen. Die hierbei ermittelten Werte für Sensitivität und Spezifität sind von den entsprechenden Werten für die Grundgesamtheit völlig und prinzipiell unterschiedlich.

[0006] Punktweise Vergleiche zwischen zwei verschiedenen Tests oder zwischen mehreren Datenerhebungen zu ein und demselben Test sind weiterhin schwierig, wenn sie nicht auf der Basis gleicher Sensitivität (bzw. Spezifität) erfolgen. Dies ist z.B. der Fall, wenn Wertepaare (Sensitivität, Spezifität) = (70%, 80%) und (50%, 90%) verglichen werden sollen. Summenmaße, in die beide Werte eingehen, wie der Youden-Index (Sensitivität + Spezifität – 1) oder die Effizienz (Anteil aller richtig bewerteten Testergebnisse an Gesamtzahl der Fälle), haben sich nicht durchgesetzt, da ihre Aussagekraft eingeschränkt ist und fehlerhafte Schlussfolgerungen möglich sind. Der McNemar-Test erlaubt also Aussagen für spezifische Wertepaare (Sensitivität/Spezifität), das heißt

für spezielle Punkte der ROC-Kurve. Die dabei vorgenommene Selektion nur eines Cut-Off-Wertes führt zu richtigen, aber eingeschränkten Aussagen.

[0007] Aufgrund dieser Nachteile ist es üblich, ROC-Analysen zur Bewertung der diagnostischen Güte vorzunehmen. Dadurch, dass diese Analysen den gesamten Wertebereich nutzen, lassen sich fehlerhafte Aussagen bei der punktweisen Analysen vermeiden. Jedoch führt die ROC-Analyse ihrem Wesen nach zu einer Übersichts-Betrachtung der diagnostischen Güte. In der Fachsprache wird dieser Sachverhalt so formuliert, dass in der ROC-Analyse jede Falsch-Positiv-Rate das gleiche Gewicht hat (Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine [Review]. Clin Chem 1993;39:561-77).

[0008] Beim Einsatz eines diagnostischen Tests interessiert den Anwender aber meist nur ein ausgewählter Sensitivitäts- oder Spezifitätsbereich. Bestehende oder fehlende Unterschiede in diesem Bereich werden durch die Gesamtbetrachtung, wie sie insbesondere durch die AUC vorgenommen wird, verdeckt. So können bestehende Unterschiede nicht als signifikant ermittelt werden. Andererseits können fehlende Unterschiede durch Auftreten von Unterschieden in anderen Sensitivitäts-/Spezifitätsbereichen fälschlicherweise als Unterschiede detektiert werden.

[0009] Der Nachteil der ROC-Analyse, der mit dem erfindungsgemäßen Verfahren insbesondere verbessert werden soll, besteht darin, dass die diagnostische Güte um den Entscheidungspunkt (Cut-Off) – die den eigentlichen Wert eines diagnostischen Verfahrens ausmacht – durch die gleichwertigen Ergebnisse beider Verfahren fern vom Entscheidungspunkt verdeckt wird.

[0010] Eine übliche Strategie zur Bewältigung dieses Nachteils ist es, nur anhand von Untergruppen die ROC-Analyse durchzuführen, wobei die Untergruppen durch einen herausgeschnittenen Wertebereich eines Markers festgelegt werden. Dieses Verfahren führt jedoch zu Verzerrungen, die eine richtige und objektive Bewertung der Unterschiede zweier Verfahren nicht zulassen. An der unteren Grenze des herausgeschnittenen Wertebereichs bestehen die Verzerrungen insbesondere darin, dass Patienten mit Erkrankung bei niedrigen Werten des zur Auswahl benutzten Markers aus der Analyse ausgeschlossen werden, obwohl sie durch diesen Marker falsch negativ und durch den zweiten Marker möglicherweise richtig positiv erkannt werden würden. Eine derartige Subgruppenanalyse mittels ROC-Kurven wird also den zur Auswahl benutzten Marker bevorzugen. Analoge Auswahlartefakte bestehen an der oberen Grenze des herausgeschnittenen Wertebereichs.

[0011] Eine andere Strategie ist die Verwendung

ausgewählter Flächenbereiche der ROC-Kurve, das heißt die Analyse von pAUC (partial Area Under Curve) (McClish DK. Analyzing a portion of the ROC curve. Med Decis Making 1989;9:190-5.). Diese Strategie führt zu richtigen Ergebnissen. Sie hat sich in der Praxis jedoch nicht durchsetzen können, da sie zu kompliziert ist, und insbesondere der statistische Nachweis von Unterschieden (Berechnung der Signifikanz und der Konfidenzintervalle) schwierig ist.

[0012] Die ROC-Analyse führt zu einer Kurve, in der die Richtig-Positiv-Rate über der Falsch-Positiv-Rate aufgetragen ist. Die entstandene Kurve ist nur für den statistisch vorgebildeten Fachmann interpretierbar. Da die ROC-Kurve häufig in medizinischen Artikeln oder Fachinformationen verwendet wird, ist eine kritische Würdigung des Potentials und der Grenzen des diagnostischen Tests durch den üblichen Anwender von diagnostischen Tests (Arzt) nicht möglich.

[0013] Weiterhin können Patienteneigenschaften wie Alter, Krankheitsstadium, Subkrankheitsklasse, die diagnostische Güte beeinflussen. Die Zusammenhänge, die zwischen der diagnostischen Güte, die z.B. anhand von Sensitivität/Spezifität Wertepaaren oder anhand einer ROC Kurve dargestellt wird, und den Patienteneigenschaften sind mit den Methoden nach dem Stand der Technik weder direkt ermittelbar noch darstellbar.

[0014] Die erfindungsgemäße Lösung zu den Problemen der aus dem Stand der Technik bekannten Verfahren besteht in dem Verfahren nach Anspruch 1 und den abhängigen Ansprüchen.

Festlegung von Cut-Off-Paaren und Ermittlung diskordanter Testergebnisse mittels Quadranten

[0015] Das erfindungsgemäße Verfahren sieht vor, dass nur die diskordanten Testergebnisse in die Analyse eingehen. Die diskordanten Testergebnisse machen den Vorteil eines diagnostischen Tests oder den Unterschied zweier diagnostischer Tests aus. Dies wird im Folgenden anhand des Vergleiches von zwei unterschiedlichen diagnostischen Tests erläutert. Für die Bewertung eines einzelnen diagnostischen Tests wird der hierfür notwendige zweite Test durch eine mathematische Beschreibung einer angenommenen, zufälligen Verteilung von Testergebnissen ersetzt.

[0016] Die diagnostischen Messwerte von zwei diagnostischen Tests, Test 1 und Test 2, werden in einem Scatterplot (Streudiagramm) aufgetragen. Durch die folgenden Verfahren können aus den so aufgetragenen Messwerten zwei Cut-Off-Werte (CO_1 , CO_2 , im weiteren Cut-Off-Paar genannt) bestimmt werden:

- Anwendung der Kriterien gleicher Sensitivität oder Spezifität.
- ein Regressionsverfahren, wie z.B. die Pas-

sing-Bablok-Regression (Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem 1983;21:709-20), in das alle Messwerte eingehen.

- ein Regressionsverfahren, wie z.B. die Passing-Bablok-Regression, in das die Messwerte nur einer der beiden zu diskriminierenden Gruppen eingehen.
- gleiche Besetzungszahlen für die Quadranten Q2 und Q4.

[0017] Durch das Cut-Off-Paar, das sich im Scatterplot durch je eine senkrechte und eine waagerechte Linie manifestiert, werden die Messergebnisse in 4 Quadranten unterteilt. Dies ist in **Fig. 1** dargestellt. Die Quadranten unterteilen die Patienten in vier Gruppen: Entweder sind beide Testergebnisse gleichartig negativ (Q1) oder positiv (Q3), oder sie unterscheiden sich (Q2 und Q4). Nach dem erfindungsgemäßen Verfahren werden nur die Patienten mit diskordantem Testergebnis einbezogen, das heißt Patienten aus den Quadranten 2 und 4.

[0018] In gleicher Weise wird über den gesamten Wertebereich des bzw. der untersuchten Tests vorgegangen. Patienten mit diskordanten Testergebnisse werden für alle, nach oben genannten Verfahren berechenbaren möglichen Cut-Off-Paare ermittelt. Wenn es demnach k verschiedene Cut-Off-Paare gibt, lassen sich jeweils k verschiedene Quadranten Q1, Q2, Q3 und Q4 angeben.

[0019] Innerhalb dieser Gesamtbetrachtung werden die Patienten für jeweils verschiedene Cut-Off-Paare meist den Quadranten Q1 oder Q3 zugeordnet. Für bestimmte Cut-Off-Paare jedoch werden die Patienten zumindest einmal entweder Quadranten Q2 oder Quadranten Q4 zugeordnet. Dabei ist es innerhalb einer Gesamtbetrachtung nicht möglich, dass ein Patient für ein Cut-Off-Paar einem Quadranten Q2 und für ein anderes Cut-Off-Paar einem Quadranten Q4 zugeordnet wird. Mit anderen Worten, alle Patienten werden eindeutig einer von zwei Subpopulationen zugeordnet.

[0020] Die Patienten der einen Subpopulation zeichnen sich dadurch aus, dass sich mindestens ein Cut-Off-Paar angeben lässt, für das er dem Quadranten Q2 zugehört. Für die andere Subpopulation gilt gleiches entsprechend für Q4. Hieraus lassen sich Charakteristiken ableiten, das heißt je Cut-Off-Paar werden für beide Subpopulationen ausgewählte Parameter bestimmt und über alle oder einen Bereich von Cut-Off-Paaren aufgetragen. Diese Charakteristiken werden im weiteren Text als Diskordanz-Analysen-Charakteristik (DAC) bezeichnet. Diese Parameter können sowohl Testergebnisse (jeweils bezogen

auf Q2 und Q4 Richtig- und Falsch-Testpositive bzw. Testnegative) sowie davon abgeleitete Größen wie Sensitivität, Spezifität, als auch Eigenschaften der Patienten, wie z.B. das Alter oder das Krankheitsstadium sein.

[0021] Auf diese Weise können in vorteilhafter Weise einerseits die diagnostische Güte selbst und andererseits auch Zusammenhänge zwischen der diagnostischen Güte und Patienteneigenschaften ermittelt und bewertet werden.

Ermittlung und Bewertung der diagnostischen Güte

[0022] Zur Ermittlung und Bewertung der diagnostischen Güte werden die Zahlen rp (richtig Testpositive), fp (falsch Testpositive), rn (richtig Testnegative) und fn (falsch Testpositive) innerhalb der diskordant eingeordneten Patienten quadrantenweise in Q2 und Q4 analysiert.

[0023] So lassen innerhalb dieser diskordant eingeordneten Patienten Sensitivitäten $rp/(rp + fn)$ und Spezifitäten $rn/(rn + fp)$ – jeweils bezogen auf die beiden diagnostischen Tests – angeben. Um Verwechslungen zu vermeiden, seien diese Parameter als DAC-Sensitivität (DAC-SENS) bzw. DAC-Spezifität (DAC-SPEC) bezeichnet. Da in den Quadranten Q2 und Q4 die Patienten die Eigenschaft haben, dass ein Testergebnis für den einen Test T1 jeweils mit dem entgegengesetzten Testergebnis für den anderen Test T2 verbunden ist, gilt:

$$rp_{T1} = fn_{T2}, rp_{T2} = fn_{T1}, rn_{T1} = fp_{T2}, rn_{T2} = fp_{T1},$$

[0024] Für Test 1 (Marker 1) gilt:

$$DAC-SENS_{T1} = rp_{T1}/(rp_{T1} + fn_{T1}) = fn_{T2}/(fn_{T2} + rp_{T2}) = 1 - rp_{T2}/(rp_{T2} + fn_{T2}) = 1 - DAC-SENS_{T2}$$

[0025] Demzufolge ergänzen sich die DAC-Sensitivitäten von Test 1 und Test 2 zu 1. Entsprechendes gilt für die DAC-Spezifitäten.

[0026] In gleicher Weise lässt sich zeigen, dass gilt: Innerhalb der diskordant eingeordneten Patienten ist der positive Vorhersagewert DAC-pVw des einen Tests gleich dem negativen Vorhersagewert DAC-nVw des anderen Tests, z.B.:

$$pVw_{T1} = rp_{T1}/(rp_{T1} + fp_{T1}) = fn_{T2}/(fn_{T2} + rn_{T2}) = nVw_{T2}$$

[0027] Wegen dieser Abhängigkeiten ist es zur Beschreibung der diagnostischen Güte ausreichend, DAC-Sensitivität (bei Verwendung von Cut-Off-Paaren nach dem Kriterium gleicher Spezifität) bzw. DAC-Spezifität (bei Verwendung von Cut-Off-Paaren nach dem Kriterium gleicher Sensitivität) nur von einem der Tests anzugeben. Ferner ist es ausreichend, entweder den positiven oder den negativen Vorher-

sagewert anzugeben.

[0028] Nach dem erfindungsgemäßen Verfahren werden derartige Parameter oder sich durch Rechenoperationen ergebende neue Parameter wie die Differenz zweier DAC-Spezifitäten in einem Graphen über den Cut-Off-Paaren oder den zur Cut-Off-Wahl verwendeten Größe aufgetragen. Zum Beispiel kann die Differenz aus der DAC-Spezifität für Test 2 und der DAC-Spezifität für Test 1 über den zusammengehörigen Cut-Off-Paaren aufgetragen werden. Wurde das Kriterium gleicher Sensitivität für die Auswahl der Cut-Off-Werte verwendet, so lässt sich diese Differenz der DAC-Spezifitäten auch über der Sensitivität auftragen. Eine auf diese Weise erhaltene Kurve hat die Eigenschaft, dass der jeweilige zu einem Cut-Off-Paar zugehörige Funktionswert nur von Patienten abhängt, die Messwerte in der lokalen Umgebung der Cut-Off-Werte aufweisen. Diese Umgebung wird von den Fällen vorgegeben, die in den zu den Cut-Off-Paaren zugehörigen Quadranten Q2, und Q4 liegen. Diese Eigenschaft ist besonders wichtig, wenn es sich um zwei Tests handelt, die stark korrelieren. Dies kann der Fall sein, wenn in den beiden Tests verschiedene Unterformen desselben Markers verwendet werden. Damit ist es im Gegensatz zur ROC-Analyse möglich, auf einfache und verzerrungsfreie Weise die diagnostische Güte für einen eingeschränkten Bereich um die Cut-Off-Werte so anzugeben, dass ein Vergleich zwischen verschiedenen Messverfahren oder zwischen verschiedenen Datenerhebungen überhaupt erst möglich wird.

Ermittlung der diagnostischen Güte im Zusammenhang mit Patienteneigenschaften (PE)

[0029] Mit dem erfindungsgemäßen Verfahren können Patienteneigenschaften PE in gleicher Weise wie die Testergebnisse einer Analyse unterzogen werden. Hierfür werden beispielsweise die Erwartungswerte der Patienteneigenschaft PE (z.B. Mittelwert des Alters) innerhalb jeweils des Quadranten Q2 und Q4 ermittelt. Dieses Vorgehen wird wieder für alle Cut-Off-Paare wiederholt, und eine Charakteristik wie oben beschrieben erstellt. Durch Gegenüberstellung der Charakteristiken für die Testergebnisse mit den Charakteristiken für die Patienteneigenschaften PE mit Hilfe von statistischen Verfahren wie z.B. Kontingenztafelanalysen, Korrelationsanalysen oder Varianzanalysen ist es möglich, Zusammenhänge der ausgewählten Patienteneigenschaften PE mit der diagnostischen Güte aufzudecken.

Analyse mit nur einem diagnostischen Test

[0030] Das erfindungsgemäße Verfahren ist auch anwendbar, wenn die diagnostische Güte nur eines diagnostischen Tests ermittelt und bewertet werden soll. In diesem Fall wird der zweite Test durch eine mathematische Beschreibung eines zufälligen Test-

ergebnisses in der gleichen Messwertverteilung wie der des ersten Tests ersetzt.

Ausführungsbeispiel

Figuren und Beispiele

[0031] Die Figuren zeigen:

[0032] [Fig. 1](#): Streudiagramm für zwei verschiedene Tests zu derselben Erkrankung

[0033] [Fig. 2](#): Auswertung der Daten Lein et al. (2003) bzgl. der Diagnostik eines Prostatakarzinoms mittels der DAC-Methode

[0034] [Fig. 3](#): Auswertung des Alters für Patienten mit Prostata-Karzinom aus der Studie Lein et al. (2003) mit der DAC-Methode

[0035] [Fig. 4](#): Auswertung der Daten aus der Studie Keller et al. (1998) zur Wertigkeit der Tumormarker CYFRA 21-1 und CEA für die Unterscheidung Plattenepithelkarzinom – Adenokarzinom

[0036] [Fig. 5](#): Auswertung der Daten aus der Studie Keller et al. (1998) zur Wertigkeit der Tumormarker CYFRA 21-1 und NSE für die Diagnostik eines Bronchialkarzinoms gegenüber benignen Erkrankungen

[0037] [Fig. 6](#): Auswertung der Tumorart für Patienten mit Bronchial-Karzinom aus der Studie Keller et al. (1998) mit der DAC-Methode

[0038] [Fig. 7a](#): Anwendung der DAC-Methode auf einen einzelnen Marker: NSE zur Diagnostik des Bronchialkarzinoms bei kleinzelligen Tumoren

[0039] [Fig. 7b](#): Anwendung der DAC-Methode auf einen einzelnen Marker: NSE zur Diagnostik des Bronchialkarzinoms bei nichtkleinzelligen Tumoren

[0040] [Fig. 2](#) zeigt ein Streudiagramm der Messergebnisse von zwei verschiedenen diagnostischen Tests, das heißt mit zwei verschiedenen diagnostischen Markern für Patienten mit (\blacktriangle) und ohne (X) Erkrankung. Durch das Einzeichnen der jeweiligen Schwellwerte (Cut-Off) CO_1 und CO_2 für jeden Marker ergeben sich vier Quadranten. Nach dem erfindungsgemäßen Verfahren werden je Cut-Off-Paar nur die Patienten eingeschlossen, die den Quadranten Q2 oder Q4 zuzuordnen sind.

Beispiel 1

[0041] Es wurde ein Vergleich der diagnostischen Güte von Gesamt-PSA (tPSA) und komplexiertem PSA (cPSA) bei der Diagnostik des Prostatakarzinoms vorgenommen.

[0042] In der Publikation Lein M, Kwiatkowski M, Semjonow A, Luboldt H-J, Hammerer P, Stephan C, et al. A multicenter clinical trial on the use of complexed prostate specific antigen in low prostate specific antigen concentrations. J Urol 2003;170:1175-9 (Lein et al., 2003) werden in einer multizentrischen Studie erhobene Daten bezüglich der Fragestellung untersucht, ob cPSA eine höhere diagnostische Güte aufweist als tPSA. In der Studie, in die nur Patienten mit einem (zuvor unabhängig gemessenen) tPSA-Messwert $< 4 \mu\text{g/L}$ einbezogen wurden, werden die Ergebnisse von ROC-Analysen wie auch Sensitivitäts-/Spezifitätswertepaare angegeben. Die Ergebnisse waren nicht eindeutig. Für den t-PSA-Wertebereich $2.5\text{--}4 \mu\text{g/L}$ wurde $\text{AUC}_{\text{cPSA}} > \text{AUC}_{\text{tPSA}}$ ermittelt. Auf der Basis von 80%, 85%, 90% und 95% Sensitivität gefundene Spezifitätsunterschiede waren dagegen nicht signifikant.

[0043] Das erfindungsgemäße Verfahren wurde auf die Daten aus Lein et al. (2003) angewendet, wobei als Parameter die Differenz der DAC-Spezifitäten verwendet wurde.

[0044] Es ergab sich, basierend auf den DAC-Spezifitäten beider Marker, die Kurve aus [Fig. 2](#) für die Differenz $\text{DAC-SPEC}_{\text{cPSA}} - \text{DAC-SPEC}_{\text{tPSA}}$.

[0045] Die beiden Abszissenachsen geben die Cut-Off-Paare an, die nach dem Kriterium gleicher Sensitivität ermittelt wurden. Ist die DAC-SPEC-Differenz > 0 , so ist cPSA überlegen, ansonsten tPSA. Die Kurve (dicke Linie) ist zusammen mit ihren punktwweisen 95%-Konfidenzgrenzen angegeben. Für die Darstellung wurde eine Glättung der Kurven vorgenommen.

[0046] Es zeigt sich, dass entsprechend der Daten Lein et al. (2003) für den tPSA-Cut-Off-Bereich von $2,5$ bis $5,5 \mu\text{g/L}$ und den zugehörigen cPSA-Cut-Off-Werten im Bereich von $2\text{--}4,5 \mu\text{g/L}$ von einer verbesserten diagnostischen Güte des cPSA ausgegangen werden kann. Es wird deutlich, dass die verschiedenen, zum Teil widersprüchlichen Aussagen in der Veröffentlichung Lein et al. (2003) durch die DAC-Methode in einer Fig. zusammengefasst werden und ihre Erklärung finden.

[0047] Das in der [Fig. 2](#) zusätzlich angezeigte Konfidenzintervall zeigt näherungsweise den Bereich an, in dem von einer signifikant besseren diagnostischen Güte des cPSA ausgegangen werden kann.

[0048] Die Ergebnisse der DAC-Methode können unter anderem dafür verwendet werden, die Festlegung von Cut-Off-Werte vorzunehmen und diese dann zu beurteilen. Es zeigt sich, dass gerade der häufig empfohlenen tPSA-Cut-Off-Wert von $4 \mu\text{g/L}$ cPSA für diagnostische Aussagen besonders geeignet ist.

Beispiel 2

[0049] Für die Patienten der in Beispiel 1 genannten Studie Lein et al. (2003) war das Alter bekannt. Für Beispiel 2 wurde das Alter in eine Auswertung entsprechend der DAC-Methode einbezogen.

[0050] [Fig. 3](#) zeigt das Ergebnis. Aufgetragen ist der zu den jeweiligen Cut-Off-Paaren (Abszissenachsen) berechnete Mittelwert der cPSA-Test-Positiven (durchgezogene Linie) und tPSA-Test-positiven (gestrichelte Linie). Es ist festzustellen, dass die durch cPSA richtig positiv erkannten PCA-Patienten im Schnitt jünger sind als die mittels tPSA richtig-positiv erkannten Patienten.

[0051] Auf der Basis dieser [Fig. 2](#) sind Rückschlüsse auf vorteilhafte Einsatzbedingungen der diagnostischen Marker, z.B. des cPSA bei jüngeren Patienten möglich.

Beispiel 3

[0052] Aus der Studie Keller T, Bitterlich N, Hilfenhaus S, Bigl H, Löser T, Leonhardt P (1998): Tumour markers in the diagnosis of bronchial carcinoma: new options using fuzzy logic based tumour marker profiles. J Cancer Res Clin Oncol 124: 565-574 (Keller et al. (1998)) zur Diagnostik des Bronchialkarzinoms mittels Tumormarkern wurden die Tumormarker CYFRA 21-1 und CEA unter dem Gesichtspunkt betrachtet, welcher Marker besser in der Lage ist, zwischen einem Plattenepithel- und einem Adenokarzinom zu unterscheiden. Bei der Analyse wurde zugrundegelegt, dass hohe CYFRA 21-1- Werte eher mit einem Plattenepithel- und hohe CEA-Werte eher mit einem Adeno-Karzinom assoziiert sind.

[0053] Die Erkennung des Adenokarzinoms wurde als „Positiv“ (Plattenepithel: „negativ“) gewertet. In [Fig. 4](#) ist die DAC-Spezifitäten für beide Marker in Abhängigkeit von der Sensitivität angegeben, die bei der Auswahl der Cut-Off-Paare verwendet wurde.

[0054] Es kann davon ausgegangen werden, dass CEA besser geeignet ist, zwischen einem Plattenepithel und einem Adenokarzinom zu differenzieren als CYFRA 21-1. Dieses Vermögen ist weitgehend gleichförmig über den gesamten möglichen Bereich von möglichen Cut-Off-Paaren gegeben.

Beispiel 4

[0055] Es wurde die diagnostische Wertigkeit von CYFRA 21-1 und NSE bei der Diagnostik des Bronchialkarzinoms untersucht. In [Fig. 5](#) ist die Differenz der DAC-Sensitivitäten für beide Marker in Abhängigkeit von der Spezifität, die bei der Auswahl der Cut-Off-Paare verwendet wurde, aufgetragen. Es ergab sich eine bessere diagnostische Wertigkeit für

CYFRA 21-1 über den gesamten Bereich der möglichen Cut-Off-Paare.

[0056] In einer zweiten DAC-Analyse wurde untersucht, inwieweit unterschiedliche Formen des Bronchialkarzinoms (nichtkleinzelliges (NSCLC) bzw. kleinzelliges (SCLC) Bronchialkarzinom) mit dem Ergebnis der besseren diagnostischen Wertigkeit für CYFRA 21-1 zusammenhängen. Dabei wurde ein Patient mit 1 kodiert, wenn er ein SCLC hatte und mit 2 im Falle eines NSCLC.

[0057] In [Fig. 6](#) sind entsprechend der DAC-Methodik die Mittelwerte dieser Kodierung für die CYFRA 21-1-Test-Positiven (durchgezogene, dicke Linie) und NSE-Test-Positiven (gestrichelte Linie) gegenüber der zu den jeweiligen Cut-Off-Paaren gehörigen Spezifität aufgetragen. Es ist festzustellen, dass die durch CYFRA 21-1 richtig positiv erkannten BCA-Patienten größtenteils NSCLC-Patienten sind, während durch NSE vorwiegend SCLC-Patienten erkannt wurden.

[0058] Aus dieser DAC-Analyse kann der Schluss gezogen werden, dass NSE und CYFRA 21-1 jeweils spezifisch eine Tumorart erkennen.

Beispiel 5

[0059] Die DAC-Methode wurde zur Bewertung der diagnostischen Güte eines einzelnen Tests angewendet.

[0060] Die diagnostische Wertigkeit des Tumormarkers NSE bei der Detektion von kleinzelligen Bronchialkarzinomen soll gegenüber der Wertigkeit von NSE bei der Detektion nichtkleinzelliger Bronchialkarzinome unter Benutzung von Daten aus der Studie Keller et al. (1998) demonstriert werden.

[0061] [Fig. 7a](#) zeigt Ergebnisse für NSE bei der Diagnostik des Bronchialkarzinoms bei kleinzelligen Tumoren und [Fig. 7b](#) zeigt Ergebnisse für NSE bei der Diagnostik des Bronchialkarzinoms bei nichtkleinzelligen Tumoren. Aufgetragen ist jeweils die Differenz des positiven Vorhersagewerts pVw des NSE zum positiven Vorhersagewert eines mathematisch geeignet beschriebenen Tests mit Zufallsergebnissen. Die Differenz ist zusammen mit ihrer unteren 95%-Konfidenzgrenze angegeben. Die Abszisse beschreibt die Sensitivität, die zur Ermittlung der Cut-Off-Paare verwendet wurde.

[0062] Es wird deutlich, dass NSE SCLC-Tumore mit besonders höher Güte ermittelt, da die pVw-Kurve in [Fig. 7a](#) deutlich höher, und über einen großen Bereich sogar bei 1 liegt.

Patentansprüche

1. Verfahren zur Ermittlung der diagnostischen Güte von diagnostischen Tests anhand von Messwerten, enthaltend die Schritte,

(a) Bereitstellung von Messwerten von zwei diagnostischen Tests (Test 1 und Test 2) gemessen an Patienten mit der (positiv) und ohne die (negativ) zu diagnostizierende Erkrankung,

(b) Festlegung von Cut-Off-Paaren $(CO_1, CO_2)_k$ nach dem Kriterium gleicher Sensitivität für die Messwerte beider Tests,

(c) für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ Auswahl aller Messwerte, die entweder

i. für Test 1 oberhalb des zugehörigen Cut-Off (CO_1) und für Test 2 unterhalb des zugehörigen Cut-Off (CO_2) liegen, oder

ii. für Test 1 unterhalb des zugehörigen Cut-Off (CO_1) und für Test 2 oberhalb des zugehörigen Cut-Off (CO_2) liegen,

(d) für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ und für die in (b) ausgewählten Messwerte Bestimmung der Anzahlen $(rp_{i,k}, fp_{i,k}, rn_{i,k}, fn_{i,k})$ von denjenigen Messwerte, die zu Patienten gehören, die richtig testpositiv $(rp_{i,k})$, falsch testpositiv $(fp_{i,k})$, richtig testnegativ $(rn_{i,k})$ und falsch testnegativ $(fn_{i,k})$ mit $i = (\text{Test1}, \text{Test2})$ getestet wurden,

(e) Berechnung von DAC-Spezifitäten $DAC-SPEC_{i,k}$ für beide diagnostischen Tests aus den Anzahlen $(rp_{i,k}, fp_{i,k}, rn_{i,k}, fn_{i,k})$ für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ nach $DAC-SPEC_{i,k} = rn_{i,k}/(rn_{i,k} + fp_{i,k})$.

2. Verfahren nach Anspruch 1, gekennzeichnet dadurch, dass die DAC-Spezifitäten $DAC-SPEC_{i,k}$ für die beiden diagnostischen Tests über den Cut-Off-Paaren $(CO_1, CO_2)_k$ oder den zugehörigen Sensitivitäten aufgetragen und verglichen werden

3. Verfahren nach Anspruch 1, gekennzeichnet dadurch, dass die Festlegung von Cut-Off-Paaren $(CO_1, CO_2)_k$ in Schritt (b) nach dem Kriterium gleicher Spezifität für die Messwerte beider Tests erfolgt und dass in Schritt (e) die Berechnung von DAC-Sensitivitäten $DAC-SENS_{i,k}$ für beide diagnostischen Tests aus den Anzahlen $(rp_{i,k}, fp_{i,k}, rn_{i,k}, fn_{i,k})$ für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ nach $DAC-SENS_{i,k} = rp_{i,k}/(rp_{i,k} + fn_{i,k})$ erfolgt.

4. Verfahren nach Anspruch 3, gekennzeichnet dadurch, dass die DAC-Sensitivitäten $DAC-SENS_{i,k}$ für die beiden diagnostischen Tests über den Cut-Off-Paaren $(CO_1, CO_2)_k$ oder den zugehörigen Spezifitäten aufgetragen und verglichen werden.

5. Verfahren nach einem der Ansprüche 1 bis 4, gekennzeichnet dadurch, dass die Cut-Off-Paare in Schritt (b) nach einem Regressionsverfahren ermittelt werden.

6. Verfahren nach einem der Ansprüche 1 bis 4,

gekennzeichnet dadurch, dass die Cut-Off-Paare in Schritt (b) so ermittelt werden, dass die Auswahl der Messwerte in Schritt (c) i. und ii. zu einer gleichen Anzahl von Messwerten in den beiden Subgruppen i. und ii. führt.

7. Verfahren nach Anspruch 1 oder 3, gekennzeichnet dadurch, dass in Schritt (e) die Berechnung der positiven Vorhersagewerte $pVw_{i,k} = rp_{i,k}/(rp_{i,k} + fp_{i,k})$ oder der negativen Vorhersagewerte $nVw_{i,k} = fn_{i,k}/(fn_{i,k} + rn_{i,k})$ aus den Anzahlen $(rp_{i,k}, fp_{i,k}, rn_{i,k}, fn_{i,k})$ für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ erfolgt.

8. Verfahren nach Anspruch 7, gekennzeichnet dadurch, dass der positive Vorhersagewert $pVw_{i,k}$ oder der negative Vorhersagewert $nVw_{i,k}$ über den Cut-Off-Paaren $(CO_1, CO_2)_k$ oder den zugehörigen Spezifitäten oder Sensitivitäten aufgetragen wird.

9. Verfahren nach einem der Anspruch 1, gekennzeichnet dadurch, dass in Schritt (e) aus den Anzahlen $(rp_{i,k}, fp_{i,k}, rn_{i,k}, fn_{i,k})$ für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ andere Parameter wie Effizienz oder Youden-Index bestimmt und über den Cut-Off-Paaren $(CO_1, CO_2)_k$ oder den zugehörigen Spezifitäten oder Sensitivitäten aufgetragen werden.

10. Verfahren nach Anspruch 1, gekennzeichnet dadurch, dass in Schritt (d) für jedes Cut-Off-Paar $(CO_1, CO_2)_k$ und für die in (c) ausgewählten Messwerte nicht die Anzahlen $(rp_{i,k}, fp_{i,k}, rn_{i,k}, fn_{i,k})$ bestimmt, sondern die Werte $PE1_{i,k}$ entsprechend einer oder mehrerer Patienteneigenschaften (PE) ermittelt werden, und dass in Schritt (e) aus diesen Werten statistische Kenngrößen wie Erwartungswert $E_{i,k}$ oder Streuungsmaß $S_{i,k}$ für die Messwerte dieser Patientengruppen berechnet werden.

11. Verfahren nach Anspruch 10, gekennzeichnet dadurch, dass die statistischen Kenngrößen über den Cut-Off-Paaren $(CO_1, CO_2)_k$ oder den zugehörigen Spezifitäten oder Sensitivitäten aufgetragen werden.

12. Verfahren nach einem der Ansprüche 1 bis 11, gekennzeichnet dadurch, dass der zweite Test in Schritt (a) durch eine mathematische Beschreibung eines zufälligen Testergebnisses in der gleichen Messwertverteilung wie der des ersten Tests ersetzt wird.

Es folgen 5 Blatt Zeichnungen

Marker2 (M2) vs. Marker1 (M1)

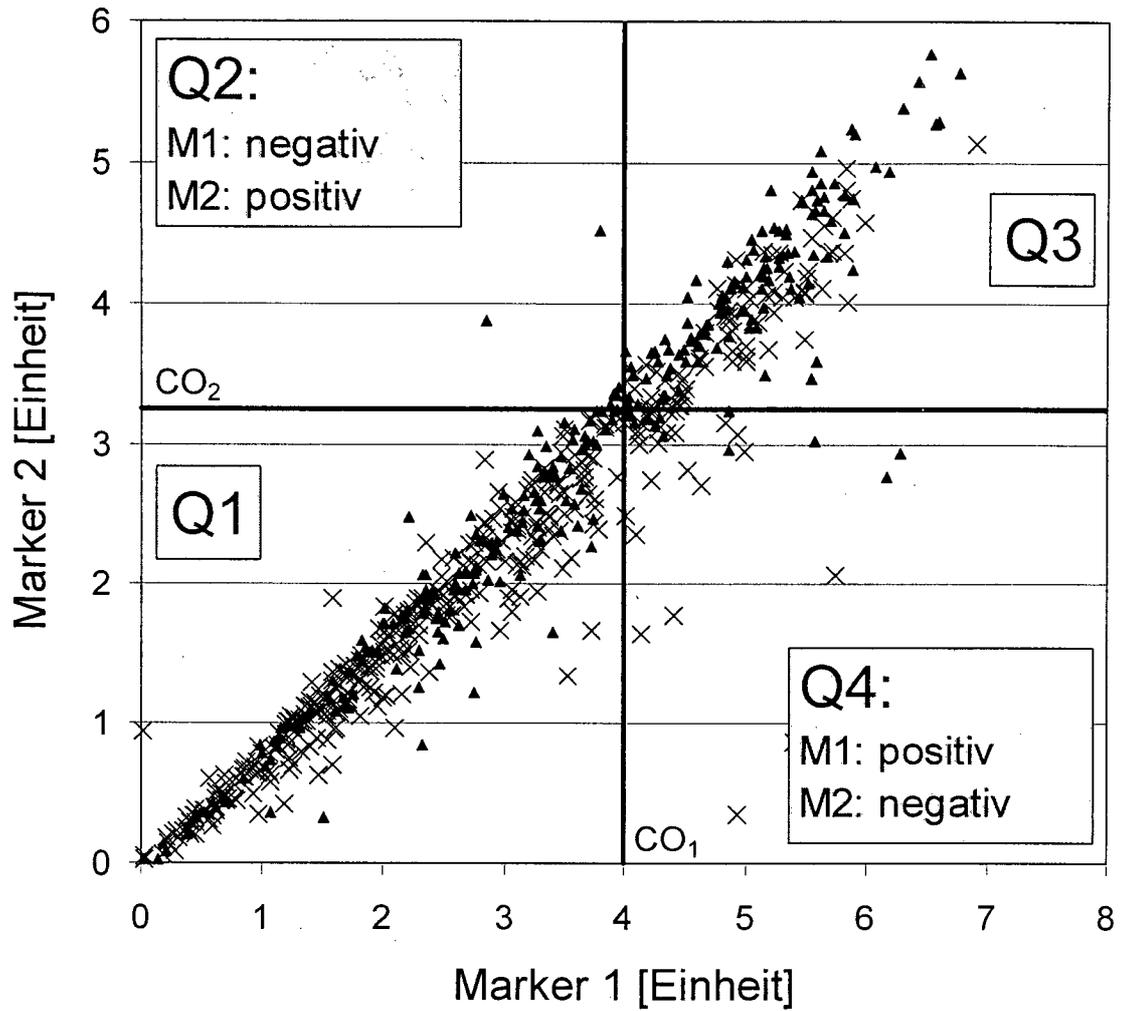


Fig. 1

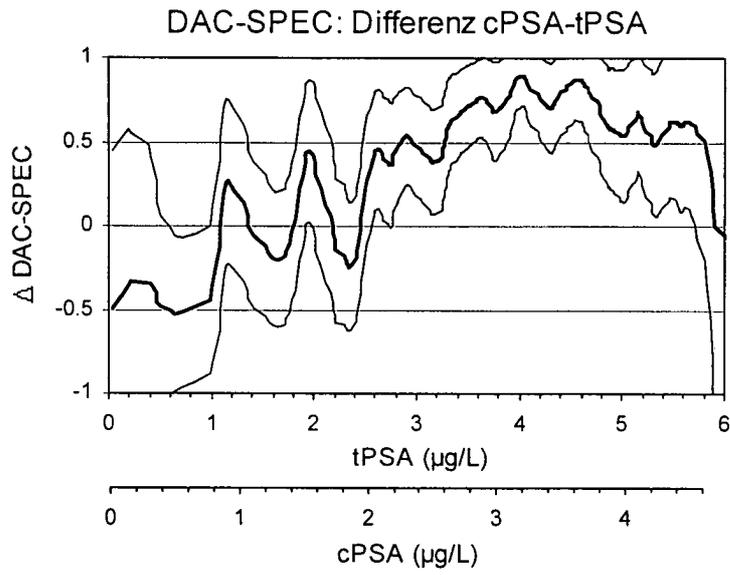


Fig. 2

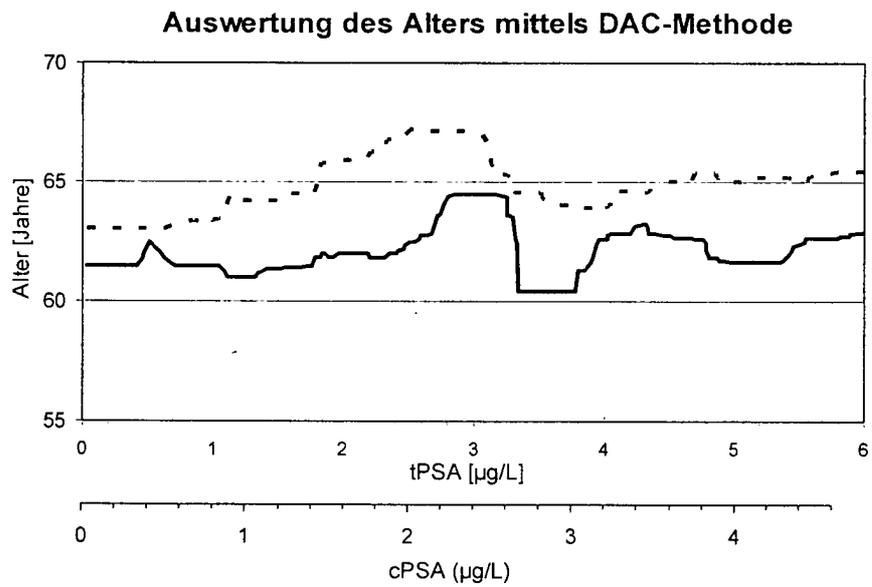


Fig. 3

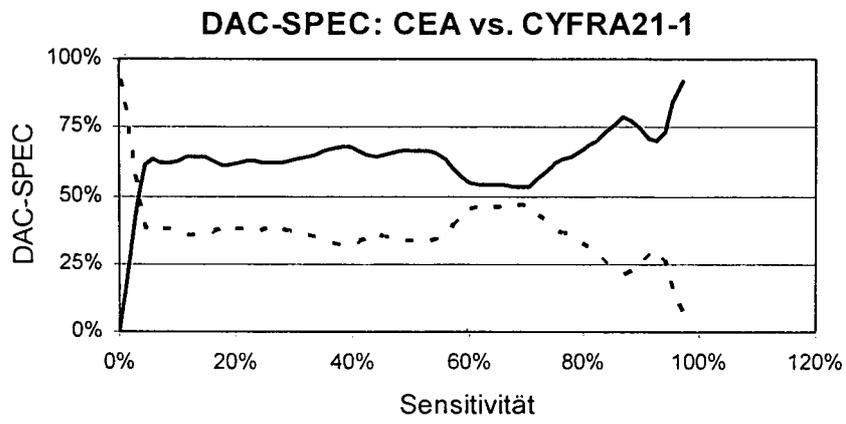


Fig. 4

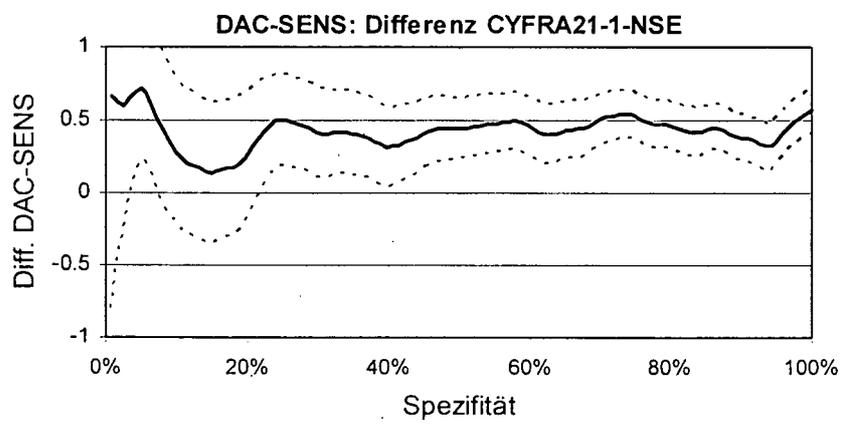


Fig. 5

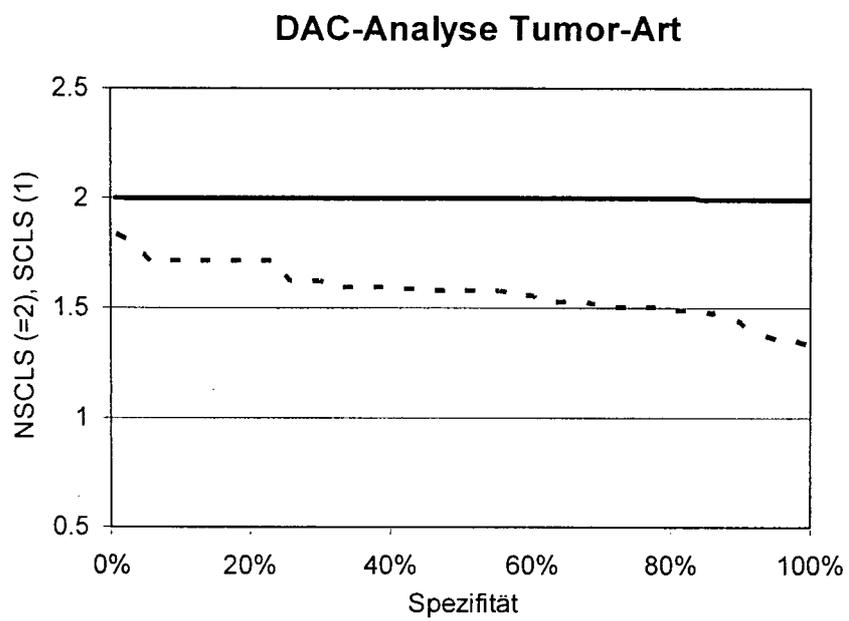


Fig. 6

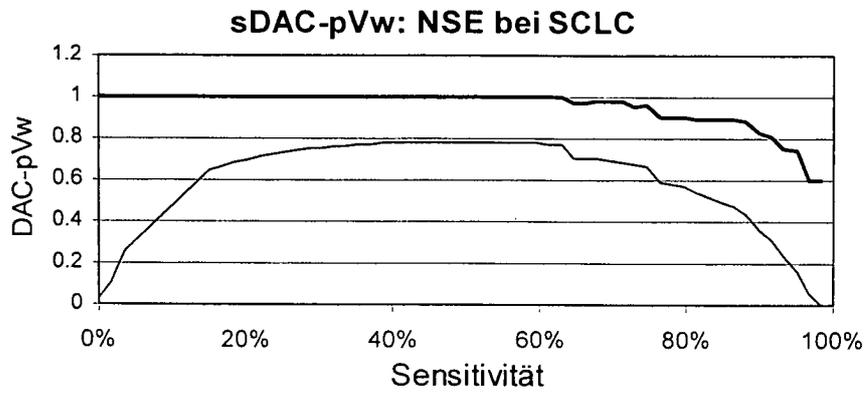


Fig. 7a

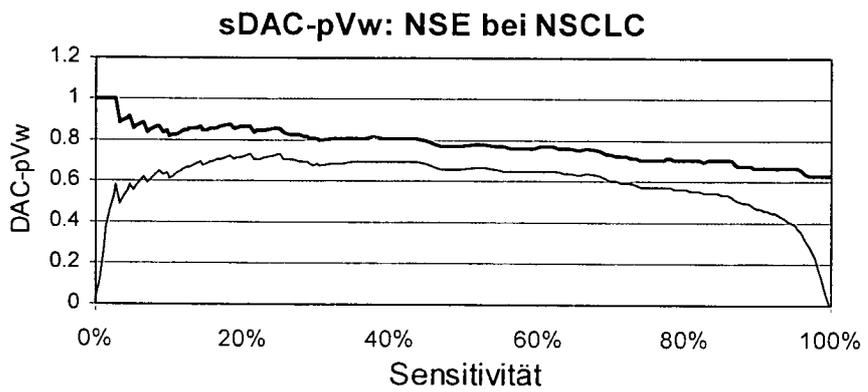


Fig. 7b