



FH Bielefeld
University of
Applied Sciences



Angewandte Mathematische Modellierung & Optimierung

AMMO – Berichte aus Forschung und Technologietransfer

Methoden der quantitativen rechnergestützten CTG-Analyse

Rainer Walden und Volker Michael Roemer

Heft Nr. 2

April 2014

Copyright

Dieses Heft wurde bereits 2009 zur Veröffentlichung eingereicht. Aus technischen Gründen erscheint diese Publikation erst 2014. Beide Autoren sind bereits nicht mehr berufstätig, aber wir wollten dem Leser diese interessante Arbeit nicht vorenthalten.

Der Herausgeber

INHALT

ABSTRACT.....	3
VORWORT.....	3
EINLEITUNG.....	4
KLASSISCHE CTG – BEWERTUNGSVERFAHREN	6
CEFMON.....	10
DEFINITION UND BERECHNUNG DES Q-SCORES.....	13
PROGNOSE DES PH-WERTS	15
DER WAS-INDEX	18
DER WAS-SCORE	23
TEST AUF AZIDOSE.....	27
ROC-ANALYSE.....	30
INFORMATIONEN	33

Abstract

This report summarises the efforts of more than ten years towards computerized CTG pattern evaluation. The results were achieved by the intensive collaboration of Prof. Roemer (former Klinikum Detmold, Detmold, Germany) and Prof. Walden (former University of Applied Sciences, Bielefeld, Germany).

This is not a medical report but deals with the technical backgrounds of the tools which were developed to achieve the results described in the papers of Prof. Roemer.

The first part shows how to standardize the notation of a score for the use in a CTG evaluation program with the aim to predict pH values in umbilical artery blood. The result is the QScore.

The second part reports about the evaluation of a new score, called WAS-Score, based on totally new ideas. This leads to an improved pH-value prediction. Moreover it gives an idea how to define a "pathological CTG".

Vorwort

Dieser Bericht gibt einen Überblick über die programmtechnischen Resultate der Zusammenarbeit der Autoren R. Walden, Prof. für Mathematik an der FH Bielefeld und Prof. Roemer, vormals Leiter der Frauenklinik am Klinikum Detmold, heute Leiter des Instituts für feto-maternale Medizin in Detmold. Es handelt sich hier nicht um einen medizinischen Bericht sondern um die im Laufe von ca. zehn Jahren mehr oder weniger intensiver Zusammenarbeit entwickelten **Modellierungen zur Bewertung der kindlichen Herzfrequenzkurven vor der Geburt durch ein Computerprogramm.**

Prof. Roemer hat das Aufkommen der Personalcomputer konsequent zur Verbesserung der wissenschaftlichen Analyse der Herzfrequenzkurven, wie sie in seiner täglichen Praxis anfielen, genutzt. Die Autoren dieses Berichts hatten also das Glück, dass sich die bei interdisziplinären Projekten üblichen Sprachschwierigkeiten in Grenzen hielten. Prof. Roemer ist der *Spiritus Rector* des Projekts, die Ideen zur Umsetzung auf dem Rechner steuerte meist Prof. Walden bei.

In den zehn Jahren der Zusammenarbeit hat Prof. Roemer ein **umfangreiches Archiv von elektronisch gespeicherten Geburtsdaten** angelegt. In dieser Arbeit wird darauf als „*das Archiv*“ Bezug genommen. Parallel zur der Erweiterung seiner Datenbank wurde ein inzwischen recht umfangreicher „Werkzeugkasten“ zu deren Analyse entwickelt. In zahlreichen Arbeiten sind die damit erzeugten Berechnungen, Bilder und Grafiken eingegangen. Die Anwendungsmöglichkeiten dieses Werkzeugkastens mag man an dem am Ende dieses Berichts aufgelisteten Verzeichnis der mit seiner Hilfe entstandenen wissenschaftlichen Arbeiten ersehen.

In letzter Zeit sind mit dem WAS-Score grundlegend neue Bewertungsmethoden erarbeitet worden, deren Tragweite noch nicht absehbar ist. Hier ist noch viel Forschungsarbeit zu leisten. Die Autoren hoffen auf das Interesse der Kollegen beiderlei Fachrichtungen. Für Interessenten finden sich die Kontaktadressen am Ende des Berichts.

Einleitung

Ein Kind kommt üblicherweise im Kreißsaal eines Krankenhauses zur Welt. Dort steht den Ärzten eine Reihe von Geräten zur Verfügung, mit denen der Zustand von Mutter und Kind vor und während der Geburt kontrolliert werden kann. Neben der Ultraschalluntersuchung, der Pulsoxymetrie und der Fetalblutanalyse (FBA) ist die Kardiotokographie (GTG) eine der Überwachungsmethoden, die üblicherweise im Kreißsaal zur Verfügung stehen.

Die CTG-Überwachung besteht in der Aufzeichnung der kindlichen Herzfrequenzkurve vor und während der Geburt, dokumentiert mit Hilfe eines Kardiotokografen, aufgezeichnet auf einem Papierstreifen (Abb.1-1) oder, moderner in Form einer elektronischen Datenaufzeichnung und Monitorwiedergabe. Ziel ist die Erkennung fetaler Gefahrendzustände, um eingreifen zu können, bevor eine Schädigung des Feten eintritt. Eine CTG-Kurve zeigt charakteristische Merkmale, die das Fachpersonal deuten kann, und die auf den Zustand des Kindes schließen lassen. Dabei steht die Erkennung von Mustern der fetalen Herzfrequenz (FHF) im Vordergrund, die mit einer mangelhaften Sauerstoffversorgung und damit Übersäuerung des Blutes des Feten verknüpft sind. Ein wesentliches Problem bei der Beurteilung eines CTG besteht darin, zu erkennen, ob eine Gefahr für das Kind vorliegt oder ob gegebenenfalls andere Gründe für die Ausprägung des CTG verantwortlich sind.

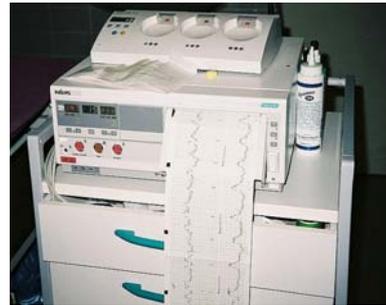


Abb. 1-1 Kardiotokograph

Quelle:

<http://de.wikipedia.org/wiki/Kardiotokografie>

Eine vermeintliche Gefahrensituation hat zu vielen unnötigen Kaiserschnitt-Geburten geführt, wie Studien in den Vereinigten Staaten gezeigt haben. Bei Tausenden von Geburten haben die Mediziner sich durch das CTG täuschen lassen. Der deutschen Gynäkologe Konrad Hammacher, des Erfinders des CTG, forderte die CTG-Papierstreifen sehr genau zu studieren um Fehlschlüsse zu vermeiden. Diese Mahnung schlugen die Geburtshelfer lange in den Wind.

Andererseits: Der umfassende Einsatz der Kardiotokographie führte zu einer Reduktion der perinatalen Sterblichkeit um den Faktor 2 und zwar sowohl bei risiko- als auch risikofreien Patientinnen (Abb. 1-2).

Die Rechtslage hat inzwischen darauf Rücksicht genommen¹:

- 1987 Oberlandesgericht Oldenburg: Eine Überwachung der fetalen Herztöne lediglich mit dem Stethoskop genügt nicht der ärztlichen Sorgfalt
- Falls nicht fortlaufend oder zumindest in Stundenabständen eine CTG - Aufzeichnung erfolgt, ist eine Beweislastumkehr gerechtfertigt.
- Ein Auskultieren der Herztöne bei Aufnahme in den Kreißsaal ist nicht ausreichend, sondern es wird eine Aufnahme - CTG gefordert

¹ Aus http://www.universitaetsfrauenklinik-tuebingen.de/uploads/media/_Ueberwachungsmethoden_unter_der_Geburt_CTG_MBU_.pdf

- Bei der Interpretation des CTG muss der Standard eines erfahrenen Facharztes sichergestellt werden. Eine Beurteilung durch die Hebamme alleine genügt nicht.

Ein weiterer Gesichtspunkt, wesentlich für die Entwicklung des automatischen Bewertungssystems, besteht in der Möglichkeit, die Stunden vor der Geburt, auch im Verbund mit anderen Überwachungsfunktionen, ohne die ständige Anwesenheit eines Spezialisten, zu überwachen und gegebenenfalls eine Warnung abzusetzen, die einen Spezialisten zu Hilfe holt. Das Fachpersonal ist natürlich nicht durch einen Rechner zu ersetzen. Aber Fachleute können nicht überall gleichzeitig sein. In diesem Sinn kann ein geeignetes Überwachungsprogramm der Entlastung des Klinikpersonals dienen.

In diesem Artikel wird über die Möglichkeiten und Methoden der elektronischen Geburtsüberwachung durch die automatische Auswertung des CTG mit Hilfe eines Rechenprogramms berichtet. Ziel der automatischen Auswertung ist es, eine Prognose über den pH-Wert des kindlichen Bluts vor der Geburt anhand der Herzfrequenzkurve zu geben. Seit den ersten Versuchen, sich von der augenoptischen Beurteilung eines CTG-Papierstreifens zu lösen und die Analyse einem PC zu überlassen, sind im Laufe der Jahre erhebliche Fortschritte gemacht worden. Dazu sei auf die entsprechenden Fachpublikationen verwiesen. In dieser Schrift stehen der algorithmische Aspekt, also das Verfahren der CTG – Bewertung, und das zugrunde liegende Entwicklungswerkzeug im Vordergrund.



Definition: Totgeborene und in der ersten Lebenswoche Gestorbene auf 1.000 Lebend- und Totgeborene.

Quelle: Statistisches Bundesamt, Statistisches Jahrbuch; Statistisches Bundesamt, Online-Veröffentlichung.

Anmerkung: Ab 2001 Westdeutschland einschließlich Gesamtberlin.

Abb. 1-2

Quelle:

http://www.gesis.org/fileadmin/upload/dienstleistung/daten/soz_indikatoren/Schluesselindikatoren/G004.pdf

Klassische CTG – Bewertungsverfahren

Üblicherweise wird der Verlauf der fetalen Herzfrequenz (FHF) über einen Zeitraum von 30 Minuten betrachtet und bewertet. Bei der Bewertung spielen die folgenden Parameter eine Rolle

1. die basale fetale Herzfrequenz (Basisfrequenz),
2. die Akzelerationen,
3. die Dezelerationen,
4. die Oszillationsfrequenz,
5. die Oszillationsamplitude (Bandbreite),
6. die Langzeitoszillationen (Oszillations-Frequenz)

Um etwa die Oszillationsfrequenz oder die Oszillationsamplitude durch einen numerischen Wert beschreiben zu können braucht man ein im Verhältnis zu 30 Minuten kleines FHF-Zeitintervall, in dem die entsprechende numerische Auswertung vorgenommen werden kann. Wir nehmen im Folgenden in solchen Fällen stets **eine Minute**. Bei einer Auswertung über 30 Minuten ergeben sich so z.B. 30 Werte für die Oszillationsfrequenz bzw. Oszillationsamplitude.

Die Definitionen solcher Parameter und ihre Bewertung erfolgt in den diversen Publikationen verbal, etwa folgendermaßen:²

Terminologie	Definition
Grundfrequenz (SpM)	ist die mittlere über mindestens 5-10 Minuten beibehaltene FHF in Abwesenheit von Akzelerationen bzw. Dezelerationen in Schlägen pro Minute (SpM) Im Bereich der fetalen Unreife liegt die mittlere FHF eher im oberen Streubereich. Ein trendmäßig zunehmender Anstieg der FHF muss besonders beachtet werden!
Normalbereich	110-150 SpM*
leichte Bradykardie	100-109 SpM
leichte Tachykardie	151-170 SpM
schwere Bradykardie	< 100 SpM
schwere Tachykardie	> 170 SpM
Bandbreite (Variabilität) (SpM)	Fluktuationen der fetalen Grundfrequenz treten 3-5mal pro Minute auf. Bandbreite ist die SpM – Differenz zwischen höchster und tiefster Fluktuation in der auffälligsten Minute innerhalb des 30-minütigen Registrierstreifens.
Normal	> 5 SpM im kontraktionsfreien Intervall

² Aus <http://www.uni-duesseldorf.de/AWMF/II/015-036.htm>

suspekt	< 5 SpM und > 40 Minuten aber < 90 Minuten oder > 25 SpM
pathologisch	< 5 SpM und > 90 Minuten
Akzelerationen	Anstieg der FHF > 15 SpM bzw. > 1/2 Bandbreite und > 15 sec
normal	2 Akzelerationen in 20 Minuten
suspekt	periodisches Auftreten mit jeder Wehe
pathologisch	keine Akzeleration > 40 Minuten (Bedeutung noch unklar)
Dezelerationen	Abfall der FHF > 15 SpM bzw. > 1/2 Bandbreite und > 15 sec
frühe	uniforme, wehenabhängig periodisch wiederholte Absenkung der FHF, früher Beginn mit der Wehe. Rückkehr zur Grundfrequenz am Ende der Wehe.
Späte	uniforme, wehenabhängig periodisch wiederholte Absenkung der FHF, Beginn zwischen Mitte und Ende der Wehe. Nadir > 20 sec nach Wehengipfel. Rückkehr zur Grundfrequenz nach dem Ende der Wehe. Bei einer Bandbreite < 5 SpM, sind auch Dezelerationen < 15 SpM gültig
variable	variabel in Form, Dauer, Tiefe und zeitlicher Abhängigkeit von Wehen, intermittierend / periodische wiederholte Absenkung der FHF mit raschem Beginn und rascher Erholung. Auch isoliertes Auftreten (in Verbindung mit Kindsbewegungen).
Atypische variable	variable Dezelerationen mit einem der zusätzlichen Merkmale: Verlust des primären bzw. sekundären FHF-Anstieges Langsame Rückkehr zur Grundfrequenz nach Kontraktionsende Verlängert erhöhte Grundfrequenz nach der Wehe Biphasische Dezeleration Oszillationsverlust während der Dezeleration Fortsetzung der Grundfrequenz auf niedrigerem Level
verlängerte	Abrupter Abfall der FHF unter die Grundfrequenz um mind. 60-90 Sekunden. Als pathologisch zu werten, wenn sie über 2 Wehen bzw. > 3 Minuten anhalten
sinusoidales Muster	Langzeitschwankung der Grundfrequenz wie Sinuswelle. Das glatte, undulierende Muster von mind. 10 Minuten, besitzt eine relativ fixe Wiederkehr von 3-5 Zyklen pro Minute und eine Amplitude von 5-15 SpM ober- und unterhalb der Grundfrequenz. Eine Grundfrequenzvariabilität lässt sich nicht nachweisen

Tabelle 1: Parameter der FHF und deren Definition (modifiziert n. RCOG und FIGO)

Quelle: <http://www.uni-duesseldorf.de/AWMF/II/015-036.htm>

Es fällt auf, dass die Oszillationsfrequenz hier nicht auftaucht. Die Beurteilungskriterien sind verbal gefasst und die Gesamtbeurteilung des CGT obliegt dem medizinischen Fachpersonal.

Die folgende Tabelle 2, die sich auf den Fischer-Score bezieht, findet sich in Wikipedia³. Die Oszillationsfrequenz wird hier als „Nulldurchgänge“ bezeichnet.

Kriterium	0 Punkte	1 Punkt	2 Punkte
Grundfrequenz	< 100/min oder > 180/min	100-110 oder 150-180	110-150
Bandbreite	< 5	5-10 oder > 30	10-30
Nulldurchgänge	< 2	2-6	> 6
Akzelerationen	keine	regelmäßige	sporadische
Dezelerationen	„ungünstige“	„günstigere“	keine

Tabelle 2: Fischer-Score

Die Beurteilung der obigen und möglicherweise weiteren Parameter wird in Form eines **Scores** ausgedrückt, also durch ganzzahlige Zahlen. Der **Gesamtscore** setzt sich aus den **Partialscores** jeder Beurteilungskategorie zusammen. Nach deren Bestimmung werden die Partialscores zum Gesamtscore addiert. Dies stellt gegenüber der Situation in Tabelle 1 einen Fortschritt in Richtung einer Formalisierung dar, wie sie für den Rechner benötigt wird. Der Gesamtscore Null stellt die letale Situation dar, in der alle Partialscores ebenfalls Null sind. Mit wachsenden Score-Werten, aufsteigend bis maximal 10, wird das wachsende Wohlbefinden des Feten beschrieben. Ein Nachteil eines solchen additiven Scores ist, dass er durchaus Werte im akzeptablen Mittelbereich annehmen kann, dass aber dabei ein Partialscore (im Extremfall) Null sein kann. Wenn z. B. die Zahl der Nulldurchgänge klein ist, kann eine Gefahrensituation vorliegen, die durch den Gesamtscore nicht ausgedrückt wird.

Die üblichen Scores sind der FIGO-Score und der Fischer-Score, die sich in den Bewertungskategorien unterscheiden. Die Bewertung ist Sache eines Spezialisten, denn die diversen Bewertungskriterien sind nicht so präzise festgelegt, dass sich stets ein eindeutiger Score-Wert ergibt. Letztlich ist hier die Erfahrung des Arztes oder der Ärztin ausschlaggebend. Es sei aber auf die in der Einleitung aufgezeigte Problematik hingewiesen.

Will man ein CTG mit Hilfe eines Computers auswerten, hat man als Grundlage und Input die FHF-Kurve. Es gilt, daraus möglichst kompetente Informationen über den Zustand des Feten zu bekommen. Informationen, die sich in der FHF-Kurve nicht niederschlagen, sind auf diese Weise natürlich auch nicht zu bekommen. Auch sind die in der FHF-Kurve steckenden Informationen in ihrer Vielfalt sicher nicht vollständig mit dem Computer erfassbar. Weiter muss man die in

³ <http://de.wikipedia.org/wiki/Fischer-Score>

allen biologischen Systemen vorzufindende Streuung der Parameter berücksichtigen, wenn man die Erfolgsaussichten eines solchen Unterfangens abschätzen will. Das Problem besteht darin, einem Computer den von einem Spezialisten augenoptisch in wenigen Sekunden erfassbaren Sachverhalt beizubringen. Dieses Problem wird natürlich nur bis zu einem gewissen Grad zu lösen sein.

Die Versuche, einen Score algorithmisch zu berechnen, also ein 30 Minuten Intervall der FHF mit Hilfe eines Rechenprogramms zu bewerten, sind zahlreich, und bisher nicht wirklich zufriedenstellend verlaufen.

In dieser summarischen Arbeit werden zwei neue Ansätze zu diesem alten Problem dargestellt, die viel versprechend sind: der **Q-Score** und der **WAS-Score**. Wissenschaftliche Details findet man in den Fachpublikationen.

Beide Scores wurden anhand eines umfangreichen **Fall-Archivs** mit über 470 Feten entwickelt, die alle auf vaginalem Weg in der Frauenklinik des Klinikums Lippe – Detmold GmbH (Leitung Prof. Dr. V. M. Roemer) 2001 bis 2008 zur Welt gekommen waren. Die elektronisch gespeicherten FHF-Kurven wurden mit einer Kopfschwartenelektrode (KSE), also direkt, abgeleitet. Alle Aufzeichnungen erfolgten mit ein und demselben Gerät.

Der Q-Score ist ein Algorithmus, dessen Erfolg u. a. auf der verfeinerten Messung der sog. Nulldurchgänge beruht, die über die augenoptischen Möglichkeiten hinausgeht und für die die Zahlenangaben in Tabelle 2 demzufolge obsolet sind. Die Zahl der „Nulldurchgänge“, genauer die Oszillationsfrequenz (ozf), gemessen durch die Zahl der lokalen Extrema der FHF-Kurve in der vergangenen Minute, ist der einzige „differentielle“ Parameter, d. h. er misst kurzfristige Veränderungen. Alle anderen Parameter sind „integraler“ Natur, d. h. die Feinstruktur der FHF-Kurve ist ohne Bedeutung. Die Oszillationsfrequenz wird im Folgenden in dieser Schrift suggestiv als **Mikrofluktuation** bezeichnet.

Es wird stets von **vier Messwerten der FHF pro Sekunde** ausgegangen, und zwar, was für die Feinstruktur wichtig ist, mit einer vom Messgerät gelieferten Genauigkeit von einer $\frac{1}{4}$ Einheit (SpM) nach dem Komma.

Natürlich müssen die Kriterien für die Vergabe der Partialscores für die oben angeführten Variablen für den Computergebrauch exakt definiert werden. Der Gesamtscore ist dann die Summe der Partialscores.

Eine solche Objektivierung der Score-Punkte-Vergabe, die in einem Rechenprogramm verwertbar ist, ist den Autoren bisher nicht bekannt. Der hier beschriebene Versuch ist also neu. Den daraus abgeleiteten wohldefinierten **quantitativen** Score nennen wir **Q-Score**.

Mit dem später zu beschreibenden **WAS-Score** wird ein gänzlich neuer Weg der CTG-Bewertung beschritten. Im Kern beruht er darauf, dass mit Hilfe der drei Parameter *Mittlere Frequenz* (f_{hm}), *Oszillationsamplitude* (o_{za}) und *Mikrofluktuation* (ozf), gemessen in jeder der 30 Bewertungsminuten, eine mathematischen Funktion aufgestellt wird, die in erstaunlich guter Weise die gemessenen pH-Werte der Fälle des Archivs prognostiziert. Der Was-Score gibt u. a. Anlass zu der Vermutung, dass sich Möglichkeiten zu einer „kontinuierlichen“ pH-Wert-Prognose vor der Geburt ergeben.

Man sollte aber stets im Auge behalten, dass mit der Reduzierung der komplexen Information eines CTG auf ein paar ganzzahlige Zahlen, sei es der FIGO-Score, der Fischer-Score oder der Q-Score, ein enormer **Informationsverlust** verbunden ist.

Auch sollte man folgenden Sachverhalt bedenken: allen statistischen Überlegungen liegt ein *Experiment* zugrunde, dessen Ausgang nach einer gewissen Verteilung dem Zufall unterliegt. Die Experimente müssen (theoretisch) beliebig wiederholbar sein, um zu statistischen Aussagen zu kommen. Das unseren Überlegungen zugrunde liegende „Experiment“ ist die Geburt eines Kindes, also nicht wiederholbar. Für uns besteht der Ausgang des Experiments in der Erzeugung einer halbstündigen FHF-Kurve vor der Geburt, die es zu bewerten gilt. Zwei gleiche Kurven bei zwei Geburten dürfte es zwar nicht geben, aber immerhin bei ähnlicher Konstellation weitgehend ähnliche Kurven. Wir gehen davon aus, dass dann auch die biochemischen Vorgänge ähnlich sind, so dass sich, innerhalb der biologischen Streuung, ähnliche pH-Werte im Blut ergeben. Der algorithmisch erzeugte Score ist dagegen nicht vom Zufall abhängig, sondern durch die FHF-Kurve eindeutig festgelegt, allerdings aufgrund des vorliegenden Datenmaterials und durch die Computer-Programme der Autoren.

CEFMON

Die in folgenden dargestellten Sachverhalte wurden mit Hilfe eines umfangreichen und über Jahre entwickelten Programm-Werkzeugs zur CTG-Analyse, genannt CEFMON (Continuous Electronic Fetal Heart Rate Monitoring) und geschrieben in Matlab⁴, gewonnen. Der berechnete Score ist der „Q-Score“.



Abb. 3-1

Mit Hilfe dieses Programms konnte die wachsende Zahl der elektronisch aufgezeichneten Fälle analysiert werden. In dem Archiv von Prof. Roemer befinden sich im Entwicklungsstadium (Juli 2009) ca. 470 CTG-Aufzeichnungen, in einem Ascii-Dateiformat, wie folgt:

Zeit	FHF
0.25	138.00
0.50	135.75
0.75	135.75
1.00	137.50
.....

⁴ Siehe www.mathworks.com

Die erste Spalte zeigt die Zeit in Sekunden, die zweite die Herzfrequenz (FHF) in Schläge pro Minute (SpM). Eine Messung besteht also aus einem Datenpaar (t_i, fhf_i) , wobei i die Werte 1, 2, 3, ... annimmt. Mit $fhf(t)$ wollen wir die Herzfrequenz zum Zeitpunkt t bezeichnen, wobei t stets einer der Zeitpunkte t_i ist. Wir werden diese Bezeichnung auch analog für andere Variable (oza , ozf , fhm , siehe unten) benutzen.

Im Augenblick der Geburt wurde in allen Fällen Blutgasanalysen im Nabelschnurblut durchgeführt, d. h. es wurden gemessen in der Arterie und der Vene: pH, pCO₂ und pO₂.

In den Studien, gewonnen mit Hilfe von CEFMON und dem Statistikprogramm JMP (Version 4, SAS-Institut, Cary, NC 27513, SAS Campus drive, USA), wurden die jeweils letzten 30 CTG-Minuten verwendet. Alle Kinder kamen auf vaginalem Weg in der Frauenklinik des Klinikum Lippe-Detmold (2001 bis 2009) zur Welt. Alle CTG's waren über Kopfschwartenelektrode (KSE), also direkt, abgeleitet worden. Das verwendete Aufzeichnungsgerät, stets ein und dasselbe HP-Gerät, lieferte vier FHF-Werte pro Sekunde mit einer Genauigkeit von $\frac{1}{4}$ Stelle nach dem Komma. Dabei taucht folgendes Problem auf:

Die Daten, die unter der Geburt vom Aufzeichnungsgerät geliefert werden, sind üblicherweise von Zeit zu Zeit durch äußere Einflüsse unterbrochen. In den aufgezeichneten Daten äußert sich das dadurch, dass in der Frequenzspalte eine **Sequenz von Nullen** erscheint. Diese Daten sind so nicht brauchbar und müssen aufbereitet werden.

Die einfachste Methode, abgesehen vom einfachen Weglassen der defekten Daten, ist eine spiegelsymmetrische Wiederholung der letzten intakten Frequenzwerte bis die Lücke geschlossen ist.

Natürlich sollte man mit einem solchen Verfahren keine großen Lücken schließen. In dem Computer-Programm CEFMON werden so nur Lücken bis zu 20 Sekunden Dauer geschlossen. Fälle mit zu großen Datenlücken müssen ausgeschlossen werden.

Die **Qualität des Q-Scores** wird in CEFMON anhand des **Korrelationskoeffizienten** zwischen den gemessenen pH-Werten pH_m und den berechneten Score-Werten Q_{sc} des Archivs beurteilt. Die numerischen Werte des Q-Score sagen wenig aus. Durch eine lineare Transformation der Q-Score-Werte kann man diese in „prognostizierte pH-Werte“ umrechnen, wozu es verschiedene Möglichkeiten gibt (siehe unten).

Für die Berechnung des Q-Scores wird die letzte halbe Stunde jedes im Archiv befindlichen Falls benötigt. Auch ohne pH-Wert kann natürlich der Q-Score einer beliebigen halben Stunde erstellt werden, zur Verwendung als „Überwachungsgröße“ während der Zeit vor der Geburt. Die Q-Score-Werte sollten dann als „prognostizierte pH-Werte“ dargestellt werden.

Dieser **Überwachungsscore** kann z. B. jede Minute neu erstellt werden, so dass sich ein quasi-kontinuierlicher zeitlicher Scoreverlauf ergibt. Das Steigen oder Fallen des Scores gibt Auskunft über das Wohlbefinden des Kindes und kann Anlass für das Eingreifen des Fachpersonals sein. Abb. 3-1 zeigt einen Überwachungsscore, der im 5-Minuten-Takt aufdatiert wurde: Die obere Kurve ist die FHF-Kurve der aktuellen 30 Minuten, darunter als rote Punkte die prognostizierten pH-Werte. Der letzte rote Punkt bezieht sich auf die aktuellen letzten 30 Minuten. Unten als Balken befinden sich die Q-Score-Werte.

Die Dezelerationen in der FHF-Kurve sind rot gekennzeichnet und haben eine Fläche von 31 Einheiten, was zu einen Partialscore von 2 führt, der Frequenzscore ist 0 und der Score der Mikrofluktuation ist 3, so dass sich ein Gesamtscore von 5 ergibt, die Balkenhöhe des letzten Balkens. Diese Score-Bewertung bezieht sich auf die Zeit von der 90ten bis zur 120ten Minute.

Eine Zusammenfassung aller Scores im 5-Minuten-Abstand zusammen mit der Prognose-pH-Werten ist im Bild 3-2 zu sehen.

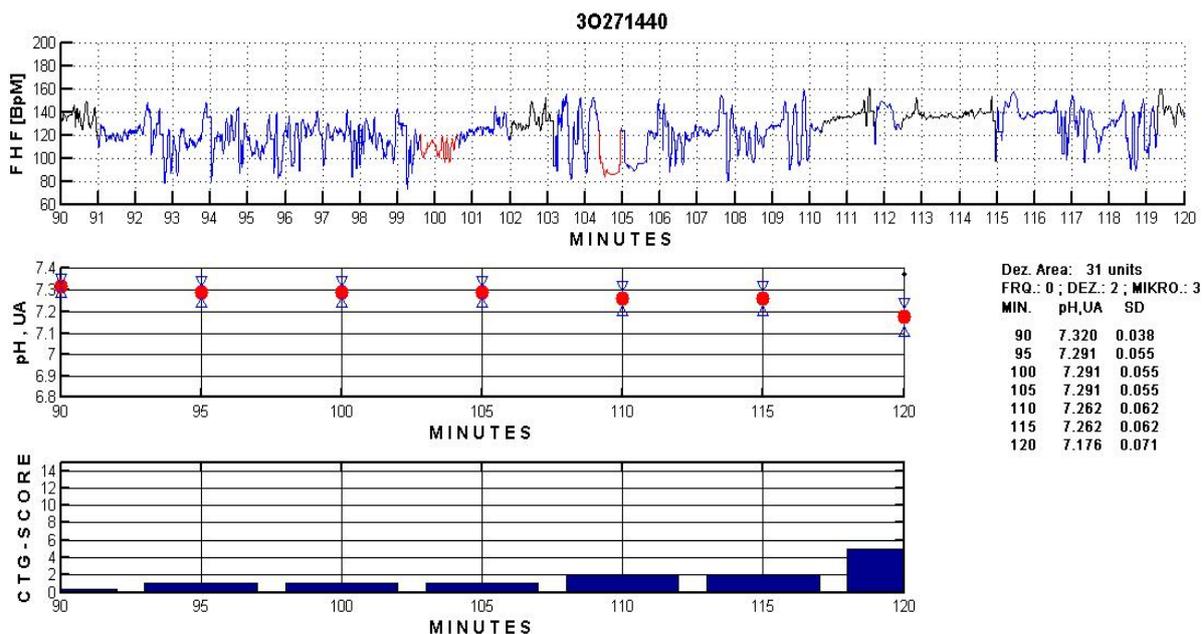


Abb. 3-2: Überwachungscore

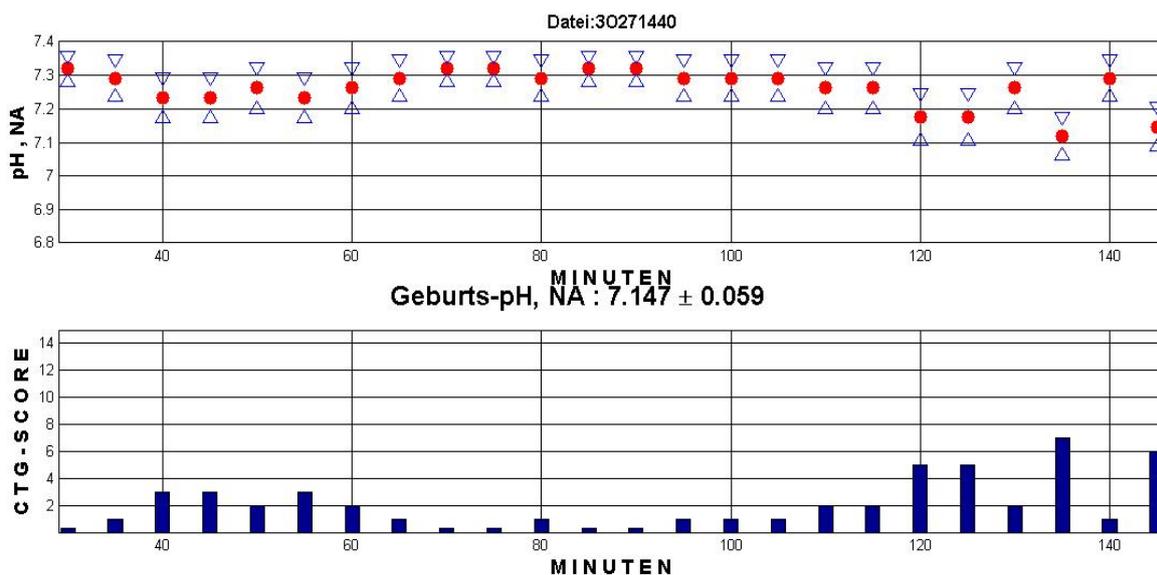


Abb. 3-3: Zusammenfassung

Definition und Berechnung des Q-Scores.

Da das „Wohl des Kindes“ keine physikalisch fassbare Größe ist, stellt sich zunächst die Frage, welchen Zweck der Q-Score dienen soll. Entsprechend kann man dann den Score entwerfen. Die Autoren haben sich entschlossen, die Score zum Zweck der pH-Wert-Prognose im Nabel-Venen-Blut zu gestalten. Dies bot sich an, da am Klinikum Detmold neben den FHF-Aufzeichnungen⁵, auch die entsprechenden pH-Wert-Messungen vorgenommen werden konnten. Im Laufe der Zeit haben sich über 470 Fälle mit mindestens 30 Minuten Aufzeichnungsdauer angesammelt, die als Grundlage der Q-Score-Entwicklung dienten.

Alle in der Literatur benannte Scores gehen von der Kenntnis der FHF-Grundlinie aus, die wir im folgenden **Baseline** nennen wollen. Optisch ist für einen Betrachter der FHF-Kurve die Baseline in den meisten Fällen recht gut zu erkennen. Die Identifizierung im Rechner macht jedoch Schwierigkeiten. Ein Algorithmus muss zu jeder FHF-Kurve eine Baseline finden. Das Problem wird in Q-Score-Algorithmus in zwei Schritten gelöst:

Im ersten Schritt wird eine *Pseudobaseline* berechnet, die sich als Linie gewichteter Mittelwert aller FHF-Werte ergibt. In einem zweiten Schritt wird festgestellt, ob und welche FHF-Abschnitte gegebenenfalls als Dezelerationen in Frage kommen. Ein Algorithmus testet Kriterien für eine Dezeleration (siehe Tabelle 1 und 2) und legt die Grenzen in Form von Anfangs- und Endpunkt fest. Dabei werden die Kandidaten für eine Dezeleration ihrer Tiefe nach abgearbeitet. Als *Baseline* wird dann der Teil der Pseudobaseline genommen, der frei von Dezelerationen ist. Abb. 4-1 zeigt ein Beispiel. Die rote Linie kennzeichnet den Bereich, in dem die FHF-Kurve zur Baseline gerechnet wird. Akzelerationen können auf analoge Weise gefunden werden, sind im Q-Score aber nicht berücksichtigt.

Das Bewertungskriterium für eine Dezeleration ist ihre Fläche, die sich als dimensionslose Größe ergibt. Die Fläche wird nach unten durch die FHF-Kurve begrenzt, nach oben durch die Strecke, die Anfangs- und Endpunkt verbindet. Die Flächen aller Dezelerationen in den 30 Test-Minuten werden addiert und in den Partial-Score-Wert umgerechnet. Liegt die Fläche zwischen 0 und 136 Einheiten, gibt es Null Punkte, zwischen 136 und 258 gibt es einen Punkt usw. gemäß der folgenden Reihe.

0...136...258...333...444...559...763...∞

Maximal sind 6 Punkte zu vergeben. Die Zahlenwerte der Reihe ergeben sich aus medizinischen Überlegungen und mit Hinblick auf eine möglichst gute Korrelation des Gesamtscores mit den gemessenen pH-Werten. Wachsende Score-Werte sollen zu fallenden pH-Werten führen.

⁵ mittels Kopfschwartenelektrode (KSE)

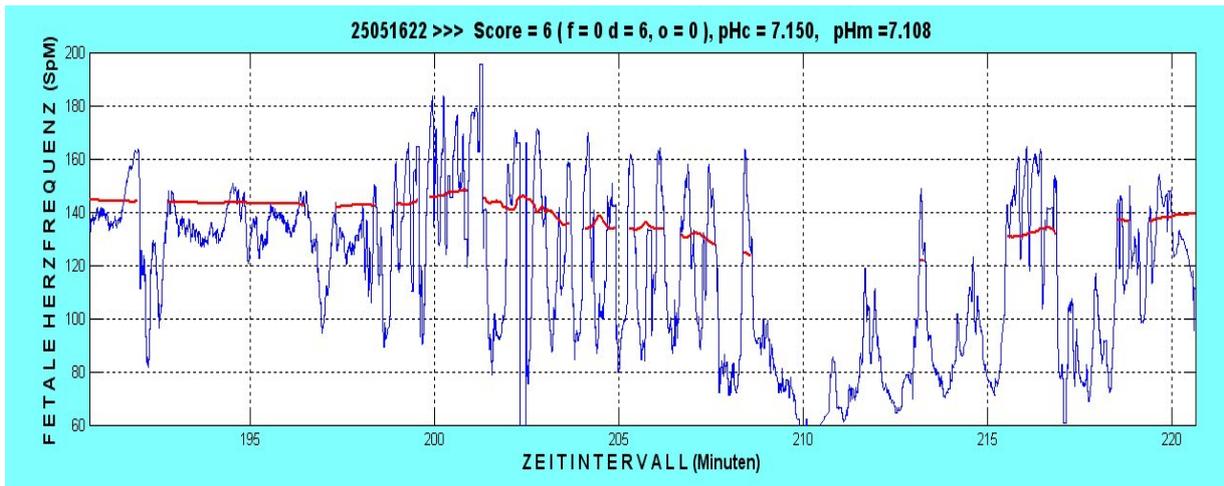


Abb. 4-1: Baseline (rot)

Anschließend werden die verbleibenden Baseline-Minuten betrachtet. Wie man aus Abb. 4-1 sieht, sind u. U. viele kleine Baseline-Stücke vorhanden, solche die länger oder kürzer als eine Minute sind. Letztere werden in der weiteren Berechnung nicht berücksichtigt. Aus den längeren werden so viele „Ganze-Minuten-Stücke“ gebildet wie möglich (und der Rest weggelassen). Für jede ganze Minute liegen 240 gemessene Frequenzwerte vor, die wir uns jeweils von $i=1$ bis $i=240$ indiziert denken. Es wird für jede dieser Minuten berechnet:

1. Die mittlere Frequenz $f_{hm} = \frac{1}{240} \sum_{i=1}^{240} fhf_i$
2. Die Oszillationsamplitude $oza = \max(fhf_i) - \min(fhf_i)$
3. Die Oszillationsfrequenz $ozf = \text{Anzahl der Umkehrpunkte}$, wobei „Umkehrpunkte“ soviel wie „lokale Extrema“ bedeutet.

Die drei Werte bilden einen Datensatz

$$(f_{hm}, oza, ozf)_j$$

für die betrachtete Minute j . j nimmt Werte zwischen 0 (keine ganze Minute in der Baseline zu finden) bis maximal 30 (die 30 Minuten bestehen nur aus Baseline) an. Es liegt also eine Verteilung dieser drei Variablen vor über die Baseline-Minuten vor. Die Score-Werte werden entsprechend dieser Verteilung vergeben. Dazu werden „Gefahren- oder Unbedenklichkeitsbereiche“ definiert, deren Grenzen durch die medizinische Erfahrung gewählt sind, und die prozentualen Anteile der in diese Bereiche fallenden Minuten bestimmt.

Intervalle für die mittlere Frequenz

0...80...171...180...∞

Liegt in der Mehrzahl (z. B. 80%) der Baseline-Minuten die mittlere Frequenz im „Normalbereich“ von 80 bis 171 Schlägen pro Minute, wird ein geringer Score-Wert vergeben, liegt ein bedenklich hoher Anteil (z.B. 50%) im Bereich von 0 bis 80 Schlägen pro Minute, wird ein hoher Score-Wert vergeben. Die Prozentwerte für die einzelnen Intervalle und die Punktevergabe sind natürlich Ermessenssache. Die Vergabe der Score-Punkte, wieder in Bereich von 0 bis maximal 6, wurde gemäß der Verteilung hauptsächlich anhand von Erfahrungswerten gewählt.

Hier wird zum ersten Mal der Versuch gemacht, dies zu objektivieren mit dem Ziel, einen möglichst hohen Korrelationswert der gemessenen pH-Werte und der Score-Werte zu erzielen. Der Weg zum Ziel besteht hier im „intelligenten probieren“.

Die „Gefahren- oder Unbedenklichkeitsbereiche“ für die Oszillationsfrequenz sind

Intervalle für die Oszillationsfrequenz **0... 20... 34... 66...∞**

Eine Scorevergabe für die Oszillationsamplitude wurde bisher nicht durchgeführt.

Die drei genannten Partialscores werden addiert, so dass sich ein Score mit Werten im Bereich von 0 bis 18 ergibt. Als übergeordnetes Bewertungskriterium, dient der Korrelationskoeffizient (aller Fälle des Archivs) des Scores S mit den gemessenen pH-Werten pH_m . Dabei wird ein möglichst hoher Korrelationskoeffizient angestrebt. Grafisch gesehen misst der Korrelationskoeffizient die Abweichung der Punkte (S_k, pH_{m_k}) (k läuft über alle Fälle des Archivs) von einer Geraden. Im Idealfall lägen alle Punkte auf einer Geraden, d. h. einem Score-Wert wäre genau ein gemessener pH-Wert zugeordnet und das Prognoseproblem wäre (zumindest für das vorliegende Archiv) gelöst. Das kann natürlich nie auftreten, wie man schon daran erkennt, dass es nur 19 Q-Score-Werte (von 0 bis 18) gibt, aber in einem Archiv mit ca. 470 Fällen natürlich mehr als 19 verschiedene gemessene pH-Werte. **Allein die „dürftige“ Punkteausstattung eines solchen Scores begrenzt also die Genauigkeit der pH-Wert-Prognose.** Im weiter unten beschriebenen WAS-Score wird dieser Mangel beseitigt.

Hier ist eine Bemerkung angebracht: als Leitlinien zur „Kalibrierung“ des Prognoseverfahrens wurde stets der Korrelationskoeffizient r_{xy} der Score-Werte x mit den gemessenen pH-Werten y betrachtet, mit dem Ziel, diesen (betragsmäßig) zu maximieren. Dieses wäre gerechtfertigt, wenn beide Variablen x und y zufallsabhängig wären. Hier ist aber nur die y-Variable (der pH-Wert) als zufallsabhängig zu betrachten. Es wäre also nahe liegender, die Fehlerquadratsumme nach Gauß zu minimieren, also eine Ausgleichsgerade von x bezüglich y zu bestimmen. Da aber r_{xy} das geometrische Mittel der Steigung der Ausgleichsgerade von x bezüglich y und der Steigung der Ausgleichsgerade von y bezüglich x ist, dürfte der Unterschied nicht von großer Bedeutung sein. In der einschlägigen medizinischen Fachliteratur ist der Begriff des Korrelationskoeffizienten besser etabliert.

Prognose des pH-Werts

Um aus den Score-Werten eine Aussage über die pH-Werte zu bekommen, machen wir eine Regressionsanalyse (einseitig, Modell I der Regressionsanalyse). Im Gegensatz zur Korrelationsanalyse (Modell II) werden hier die Einflussgröße (Regressor) X als fest angesehen und nur die Zielgröße (Regressand) Y als zufällig betrachtet. Wir betrachten also die Score-Werte 0 bis 18 nicht als zufällige Größen, sondern fest vorgegeben.

In der Abb. 5-1 sind die gemessenen pH-Werte über den zugehörigen berechneten Score-Werten für jeden Fall aufgetragen. Die „Länge“ der „Streifen“ in der Grafik geben einen Eindruck von der Variabilität der pH-Werte zu einem bestimmten Score. Es ist ferner eine Ausgleichsgerade und eine Ausgleichsparabel

eingetragen. Beide Linien sind nach der Methode der kleinsten Fehlerquadrate (nach Gauß) berechnet.

Wir beschränken uns auf die Ausgleichsgerade $y = ax + b$ und ordnen jedem Q-Score-Wert als Prognosewert den pH-Wert zu, der auf der Geraden zu finden ist

$$pH_{prog} = a \cdot QScore + b$$

Der prognostizierte pH-Wert entsteht also durch eine lineare Transformation der Q-Score-Werte, was bedeutet dass die Korrelationskoeffizienten $pH_m - QScore$ und $pH_m - pH_{prog}$ gleich sind. In Abb. 5-2 sehen wir das Resultat dieses Verfahrens. Die Regressionsgerade ist eingezeichnet.

Ein wenig Statistik: Führt man eine Korrelationsanalyse durch, so erhält man, mit der Bezeichnung X für den Q-Score und Y für die pH-Werte,

Mittelwert	$X = 3.5354$	Mittelwert	$Y = 7.2617$
Variation	$X = 3.4852$	Variation	$Y = 0.0042651$
Streuung der Prognose	0.0030942		
Bestimmtheit	0.27604		
Kovarianz	-0.064057		
Korrelationskoeffizient	0.5254		

Der Test auf Unabhängigkeit von X und Y wird mit einer Irrtumswahrscheinlichkeit von 0.01 abgelehnt. Das Konfidenzintervall für die Steigung der Geraden ist $[-0.01871, -0.018049]$.

Der Q-Score in der beschriebenen Weise ist als Modul bei den Überwachungsgeräten der Firma Nexus/GMT⁶ realisiert.

Zu bemerken ist, dass es keineswegs zwangsläufig ist, die Ausgleichsgerade zur Umrechnung von Q-Scores in pH-Werte zu benutzen. Zwar wird bei diesem Verfahren nach Gauß die Summe der Fehlerquadrate minimiert. Sinnvoll wäre aber z. B. auch eine Approximation nach Tschebyscheff, bei der der maximale Fehler minimiert wird. Außerdem könnte man bei einer Realisierung des Verfahrens die Prognose auf den kritischen Bereich der niedrigen pH-Werte konzentrieren.

Ein Problem ist dabei natürlich die geringe Fallzahl der Azidosen.

⁶ <http://www.nexus-ag.de>

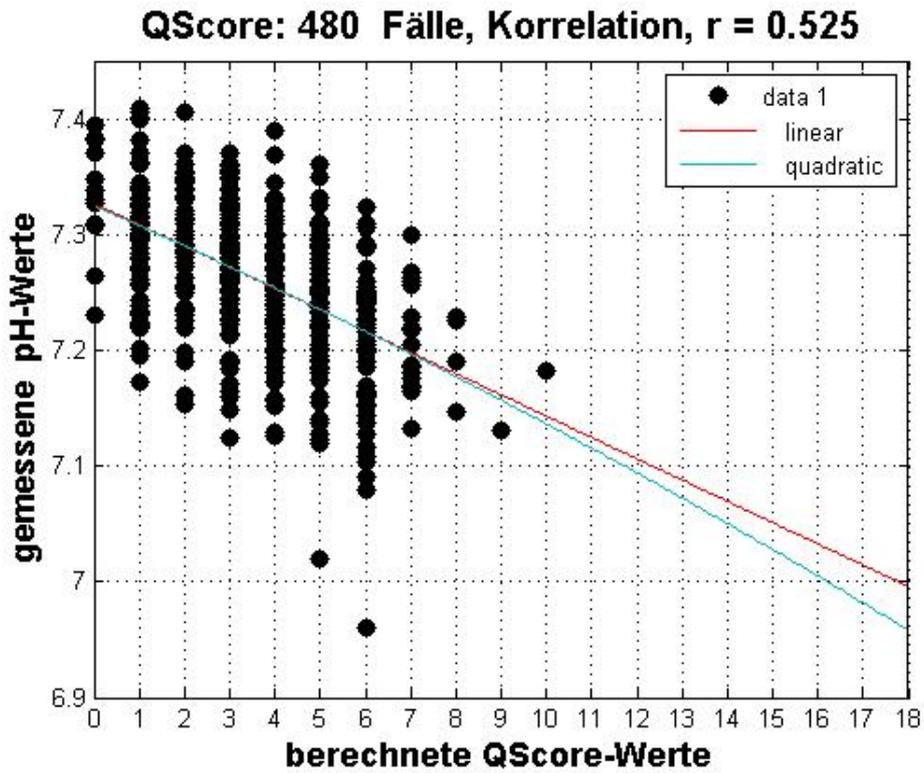


Abb. 5-1

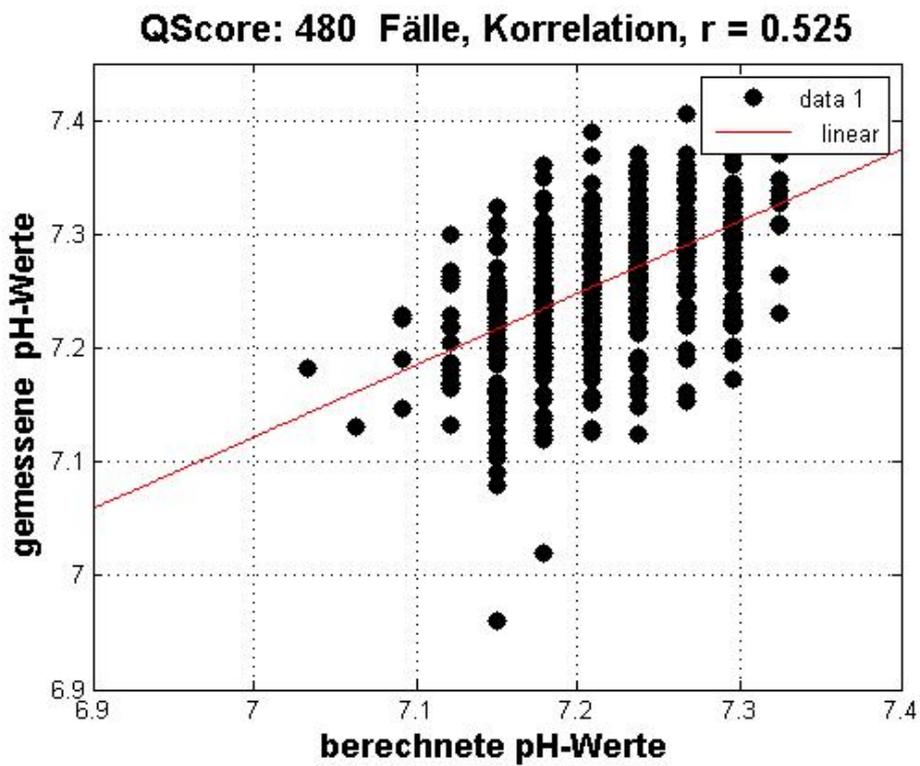


Abb. 5-2

Der WAS-Index

Wir beschreiben zunächst den WAS-Index, aus dem der WAS-Score abgeleitet wird. Als **Index** bezeichnen wir generell einen Zahlenwert, der aus einem kurzen Stück der FHF-Kurve gewonnen wurde, üblicherweise aus wenigen Minuten. Ein **Score** dagegen wird stets aus einer 30 Minuten langen FHF-Kurve berechnet. Zeitlich gesehen ist also ein Index ein „lokaler“ Wert, ein Score ein „globaler“.

Die Entwicklung des Q-Scores hat gezeigt, dass ein optisch nicht erkennbarer Sachverhalt, nämlich die Mikrofluktuation, eine wesentliche Rolle spielt. Ein Stück der FHF-Kurve, das optisch schwache Dynamik zeigt, und somit als Gefahrensignal verstanden werden könnte, kann durchaus auf verschiedene Weise zustande gekommen sein. Streckenweise ist optisch eine Azidose nicht oder nur schwer von einem schlafenden Feten zu unterscheiden. Die Untersuchungen zum Q-Score haben gezeigt, dass das Verhältnis von Oszillationsfrequenz und Oszillationsamplitude ein wichtiger Indikator für den Zustand des Feten ist.

Wachzustand, Schlafzustand und Azidosegefahr sollten vom Computer unterschieden werden können. Allerdings verlassen wir mit solchen Überlegungen die klassische Methode der Score-Bewertung, bei der standardmäßig ein 30 Minutenintervall beurteilt wird, wobei z.B. die Reihenfolge des Auftretens von Dezelerationen oder der Reihenfolge, in der die Baseline-Minuten ausgewertet werden, keine Rolle spielt. Wenn wir einen zeitlich veränderlichen Zustand eines Feten wie Schlafen und Wachen erkennen wollen, müssen wir den zeitlichen Verlauf der FHF-Kurve auch in kleineren Intervallen als 30 Minuten untersuchen. Dies führt zum Begriff des **Was-Index (Wachen-Azidose-Schlafen)**.

Aus der Erkenntnis heraus, dass das Verhältnis von Mikrofluktuation (ozf) zu Oszillationsamplitude (oza) von gravierender Bedeutung ist, und möglicherweise der Mittelwert der Frequenz (fhm) (gebildet über dieselbe Minute, die auch zur Berechnung von ozf und oza benutzt wird) eine Rolle spielt, wurde der Ausdruck

$$\frac{ozf \cdot fhm}{oza}$$

untersucht. Zu seiner Berechnung ist stets eine FHF-Minute notwendig, und der Wert dieses Ausdrucks wird dem Endzeitpunkt t dieses Minuten-Intervalls zugeschrieben. Wir schreiben also

$$W(t) = \frac{ozf(t) \cdot fhm(t)}{oza(t)},$$

wobei zur Berechnung das Zeitintervall $[t-1, t]$ benötigt wird (wenn die Zeiteinheit die Minute ist).

Es ist klar, dass dieser Ausdruck eigentlich nur eine Funktion von drei Variablen ist.

Wir suchen eine Funktion so, dass ein großer Wert auf einen gesunden Feten hinweist, und ein gegen Null gehender Wert wachsende Gefahr anzeigt.

Dazu ist z. B. bei der mittleren Frequenz fhm folgendes zu tun: ähnlich wie in Tabelle 2 (Fischer-Score) muss der Wert der Funktion sinken, wenn das Wohlfühl-Frequenzband (2 Punkte) verlassen wird. Das kann man durch eine *Gewichtung* der Variablen fhm erreichen, die die Score-Verteilung nachahmt. Die Gewichtung könnte z. B. im „Wohlfühlbereich“ 1 sein und in den „Gefahrenbereichen“ absinken (sogar auf 0, z. B. wenn fhm gegen 0 SpM oder 200 SpM gehen).

Die folgende Tabelle 3 zeigt den Schwankungsbereich der Variablen, ausgewertet über 14100 sich nicht überdeckende Minuten (jeweils die letzten 30 Minuten von 474 Fällen bei maximal zulässiger Fehlerrate von 6,7% und Datenspiegelung als Reparaturverfahren).

	Min	Mean	Median	Max
<i>fhm</i>	50 SpM	128 SpM	129 SpM	210 SpM
<i>ozf</i>	0 pM	52 pM	52 pM	123 pM
<i>oza</i>	0 SpM	46 SpM	42 SpM	179 SpM

Tabelle 3: Statistik aller letzten 30 Minuten von 474 Fällen

Man muss also z. B. die mittlere Frequenz *fhm* mindesten in Bereich von 50 SpM bis 210 SpM gewichten, man könnte auch sagen: mit einem kontinuierlichen Score versehen. Wir schreiben

$$FRQ = fhm \cdot g_{fhm}(fhm),$$

wobei $g_{fhm}: [50, 210] \rightarrow \mathbb{R}$ eine reellwertige Funktion ist. Sie wird in der Nähe von 128 SpM ein Maximum haben und nach rechts und links absinken.

Analog muss man natürlich bei den anderen Variablen *oza* und *ozf* vorgehen:

$$AMP = oza \cdot g_{oza}(oza)$$

$$MIKRO = ozf \cdot g_{ozf}(ozf)$$

Aus diesen drei Funktionen kann man Funktionen

$$W(FRQ, AMP, MIKRO)$$

bilden. Von Interesse sind natürlich nur solche, die einen Bezug zur Befindlichkeit des Kindes haben. Trotz dieser Einschränkung dürften die Möglichkeiten zahllos sein.

Wir haben uns nach einigen Tests für folgendes entschieden. Den Wert

$$W(t) = \frac{MIKRO(t) \cdot FRQ(t)}{AMP(t)}$$

bezeichnen wir als **Was-Index**. Er kann berechnet werden, wenn die FHF-Kurve der Minute von $t-1$ bis t bekannt ist.

Das Problem mit dem WAS-Index liegt natürlich in der Auswahl der Gewichtsfunktionen, für die ein Mensch kaum ein Gefühl hat. Wie kann man

$$g_{ozf}(ozf), g_{oza}(oza) \text{ und } g_{fhm}(fhm)$$

bestimmen? Die Gewichtsfunktionen haben die Aufgabe, bestimmte Wertebereiche der Variablen *ozf*, *oza* und *fhm* in ihrer Bedeutung zu stärken oder schwächen, wie dies bei der Punktevergabe der klassischen Scores auch geschieht.

Als Gewichtsfunktionen benutzen wir stückweise lineare Funktionen. Eine stückweise lineare Funktion ist eine Funktion, die durch wenige **Stützstellen** x_1, x_2, \dots (auf der x-Achse) und zugehörige **Stützwerte** y_1, y_2, \dots (auf der y-Achse) dadurch festgelegt ist, dass die Zwischenwerte zwischen diesen Werten durch lineare Interpolation gewonnen werden. Grafisch ist eine stückweise lineare Funktion ein Polygonzug durch die **Stützpunkte** $(x_1, y_1), (x_2, y_2), \dots$.

Die Gewichtsfunktionen in CEFMON benutzen die Stützstellen

0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000

für die Amplitudenfunktion und die Mikrofluktuation, und

0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 1000

Die Werte der Gewichtsfunktion an diesen Stützstellen zu setzen, also die Festlegung der Stützwerte, ist selbst mit medizinischen Erfahrung eine schwierige Sache.

Die folgende Tabelle zeigt die von den Autoren (nicht persönlich) gewählten Werte der Gewichtsfunktion $g_{oza}(oza)$ bei 40 SpM, bei 50 SpM und 60 SpM. Die Zwischenwerte wurden durch Interpolation gewonnen.

oza	goza	
40.0000	0.1679	← vorgegebene Gewichtung bei 40 SpM
41.0000	0.1652	
42.0000	0.1626	
43.0000	0.1599	
44.0000	0.1573	
45.0000	0.1546	
46.0000	0.1520	
47.0000	0.1493	
48.0000	0.1467	
49.0000	0.1440	
50.0000	0.1414	← vorgegebene Gewichtung bei 50 SpM
51.0000	0.1420	
52.0000	0.1427	
53.0000	0.1433	
54.0000	0.1440	
55.0000	0.1446	
56.0000	0.1453	
57.0000	0.1459	
58.0000	0.1466	
59.0000	0.1472	
60.0000	0.1479	← vorgegebene Gewichtung bei 60 SpM

Wie die Wahl der Stützpunkte zustande kommt, beschreiben wir in Zusammenhang mit den WAS-Score (siehe unten).

Der WAS-Index ist zu jedem Zeitpunkt der FHF-Kurve berechenbar und benötigt dazu die FHF-Werte der vorangegangenen Minute. Somit kann man den Was-Index praktisch kontinuierlich (d.h. jede ¼ Sekunde) berechnen. Um den Rechenaufwand in Grenzen zu halten wird in Programm CEFMON der Was-Index nur jede Sekunde berechnet und die Zwischenwerte werden interpoliert.

Der Was-Index ist ein sehr virulentes Gebilde, wie man an in Abb.6-1 sieht. Die blaue Kurve ist die FHF-Kurve, die rote Kurve ist der Was-Index. Der Was-Index selbst hat numerische Werte, die durch eine lineare Transformation verändert werden können, ohne dass sich der Informationsgehalt der Kurve ändert. Wir haben hier eine Transformation benutzt, die für den WAS-Score entwickelt wurde (unten). Diese Transformation rechnet WAS-Score-Werte in pH-Werte um, nach dem gleichen Prinzip wie beim Q-Score.

Hier zeigt sich das Phänomen, dass der Was-Index im Zusammenhang mit einer „simultanen pH-Wert-Prognose“ stehen könnte. Ein statistischer Beweis dafür steht natürlich aus (Messverfahren?). Die horizontale Linie unten ist der Mittelwert aller gemessenen nichtazidotischen pH-Werte bei 7,25. Die blauen Punkte sind der (unten beschriebene) WAS-Score.

Um die Virulenz des Was-Index zu dämpfen, greifen wir zu einer „gleitenden Mittelwertbildung“ oder „Glättung“. Die Glättung bezieht sich immer auf die Werte des Was-Index eines unmittelbar zurück liegenden Zeitintervalls.

Bei einer Glättung über 3 Minuten ergibt sich ein deutlich ruhigeres Bild (Abb. 6-2). Man kann hier schon besser von einer „simultanen pH-Wert-Prognose“ sprechen.

Übertreibt man es mit der Glättung, so erhält man eine fast horizontale Linie. In Abb. 6-3 haben wir eine Glättung über 30 Minuten. Dies kann man als Mittelwertbildung des Was-Index über die letzten 7200 Werte ansehen.

Den (über eine gewisse Zeitspanne) geglätteten Was-Index wollen wir der Einfachheit halber auch **Was-Index** nennen und mit $W_{index}(t)$ bezeichnen, ohne jedes Mal auf das Glättungsintervall hinzuweisen. Der Was-Index $W(t)$ stimmt also mit dem Was-Index $W_{index}(t)$ überein, wenn keine Glättung vorgenommen wird.

Ein Glättungsintervall von 3 bis 5 Minuten hat sich als sinnvoll erwiesen. Wir schreiben also

$$W_{index}(t) = \frac{1}{n} \sum_{k=t-n}^t W(t_k)$$

Hier läuft k über alle FHF-Zeitpunkte von $k=t$ rückwärts bis $k=t-n$. Für 3 Minuten sind das z.B. $n=3*60*4=720$ Zeitpunkte. An allen diesen Zeitpunkten werden die W -Werte bestimmt, (zur vollen Sekunde gemäß der Formel, dazwischen durch Interpolation) und dann der Mittelwert gebildet.

Mit Hilfe des Was-Index ist es möglich, den Zustand des Feten genauer zu diagnostizieren (Schlafen, Wachen, Azidose).

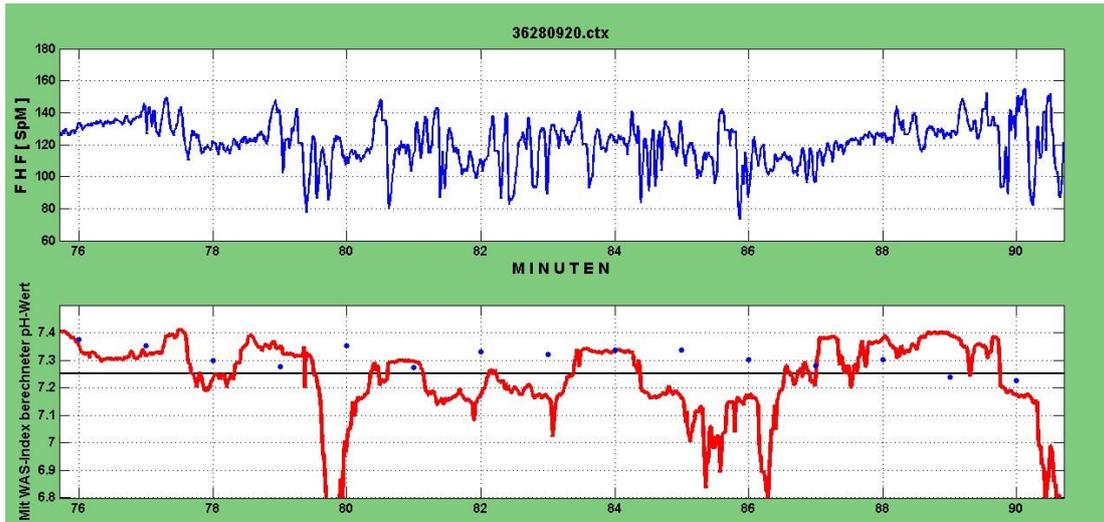


Abb. 6-1

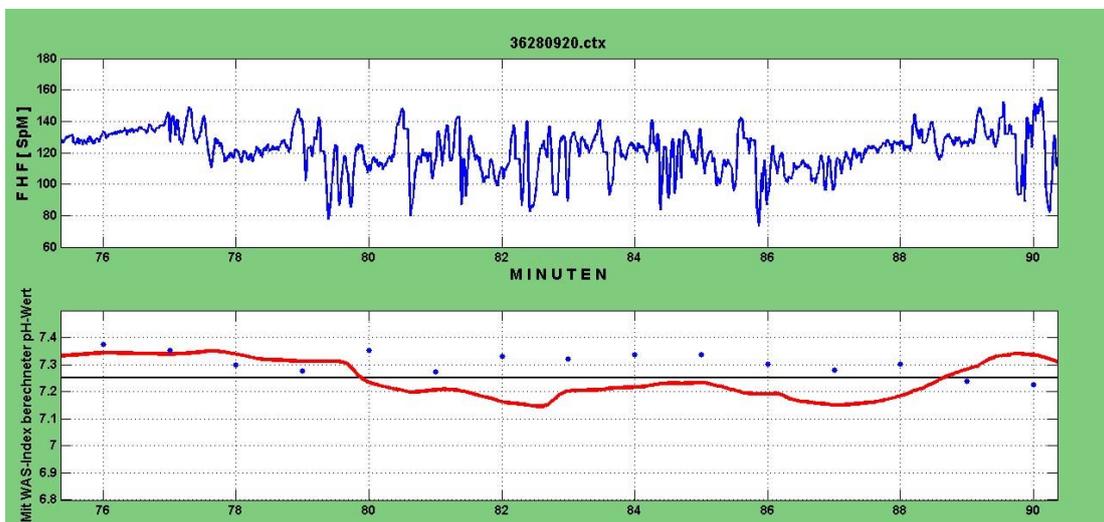


Abb. 6-2

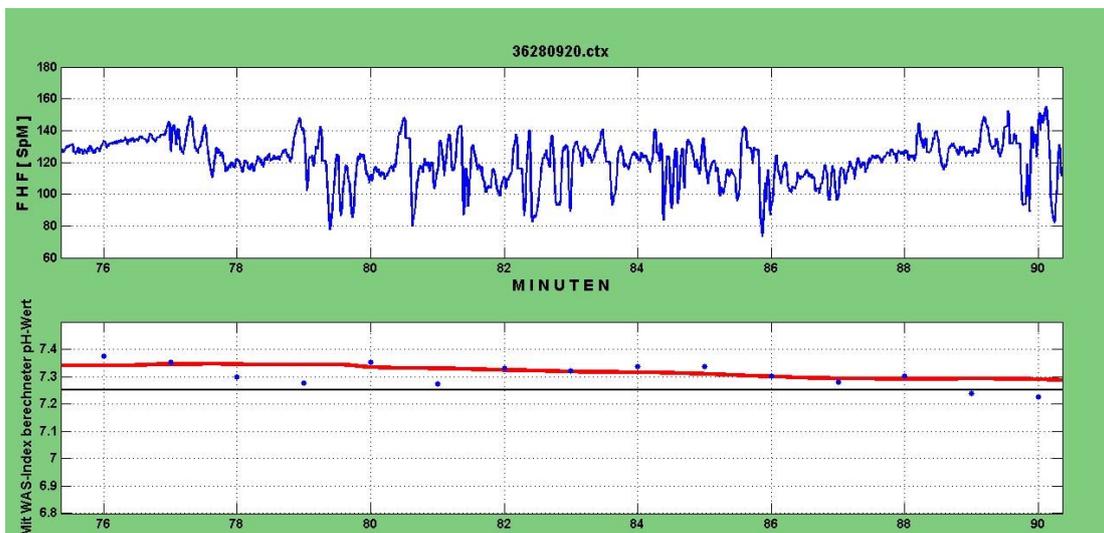


Abb. 6-3

Der WAS-Score

Der Q-Score ist ein **additiver** Score mit wenigen **diskreten** möglichen Werten und beruht, wie alle klassischen Scores, auf einer Bewertung der optisch erkennbaren Variablen Basisfrequenz, Akzelerationen, Dezelerationen, Oszillationsfrequenz, Oszillationsamplitude, und gegebenenfalls weiterer optisch erkennbarer Phänomene.

Der Was-Score ist ein **multiplikativer** Score mit einem **Kontinuum** an möglichen Score-Werten und wird nur aus den drei aus der FHF-Kurve abgeleiteten Größen mittlere Frequenz (*fhm*), Oszillationsamplitude (*oza*) und Mikrofluktuation (*ozf*) gebildet. Eine Unterscheidung zwischen Baseline und Nichtbaseline gibt es nicht.

Der WAS-Score dient dazu, 30 Minuten eines FHF-Datensatzes durch eine Zahl zu bewerten, die Information über den Zustand des Feten geben soll. Wir definieren in Anlehnung an den Was-Index den **WAS-Score** wie folgt:

$$W_{Score} = \frac{1}{30} \sum_{k=1}^{30} W(t_k)$$

Die Zeitpunkte t_k sind die Minuten 1 bis 30 des Datensatzes.

Der WAS-Score kann für jede halbe Stunde berechnet werden. In den Abbildungen 6-1 bis 6-3 sind die WAS-Score-Werte, umgerechnet in pH-Werte, als blaue Punkte dargestellt. Ein blauer Punkt ist also der WAS-Score umgerechnet in einen pH-Wert, berechnet aus den Daten der letzten 30 vergangenen Minuten.

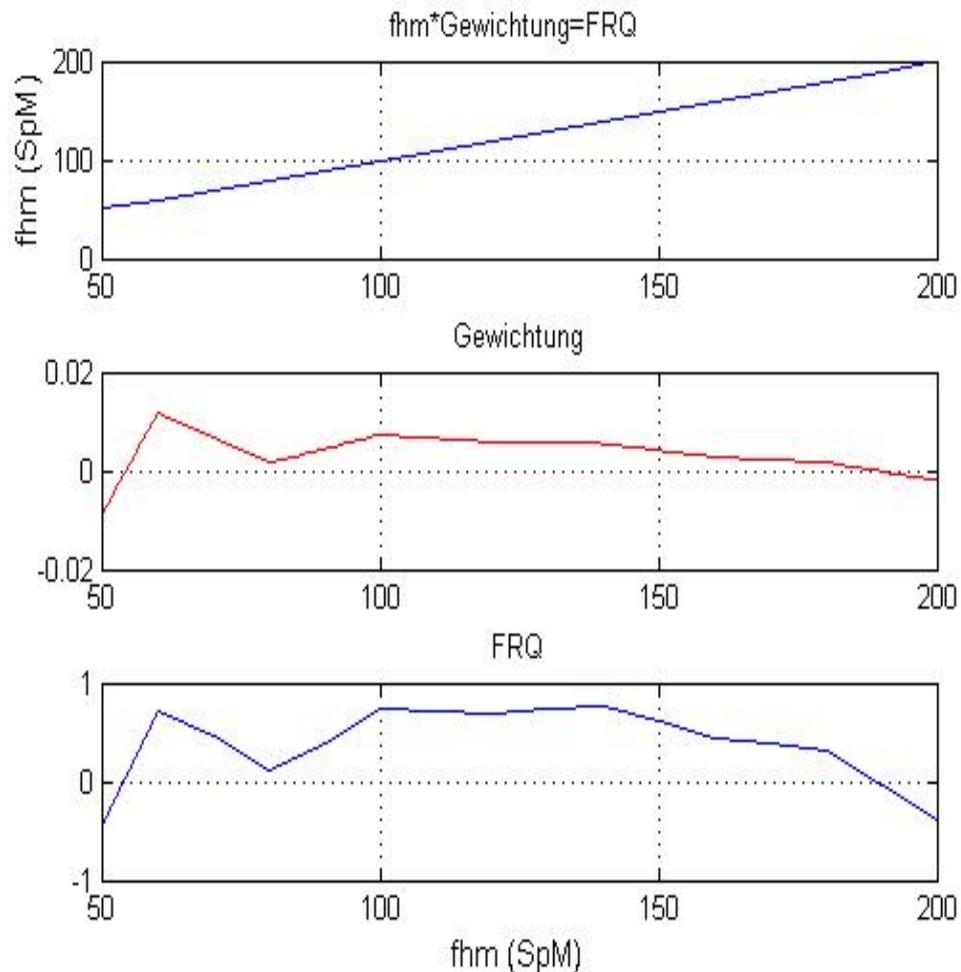
Um einer besseren Übersichtlichkeit willen transformieren wir die WAS-Index-Werte unseres Archivs durch eine lineare Transformation (Skalierung) auf das Intervall von Null bis Eins, wobei 0 für den kleinsten und 1 für den größten Wert steht. Dennoch kann ein neu hinzugekommener Fall natürlich einen (geringfügig) unter Null oder über Eins liegenden WAS-Score haben, was eine Neuskalierung ratsam machen würde. Eine Folge dieser Skalierung ist, dass auch die WAS-Score-Werte zwischen 0 und 1 liegen.

Wir kommen jetzt zu einem zentralen Problem: Wie sollen die Gewichtsfunktionen festgelegt werden. Wir müssen „passende“ Stützwerte für unsere drei Variablen *fhm*, *oza*, *ozf* an den Stützstellen wählen.

Wir gehen wie beim Q-Score vor. Im Archiv befinden sich über 470 FHF-Datensätze mit gemessenen pH-Werten. Ziel der Gewichtung ist es, eine möglichst hohe Korrelation zwischen dem WAS-Score und den gemessenen pH-Werten zu erreichen. Bei dieser **Optimierung des Korrelationskoeffizienten** stößt die medizinische Erfahrung bald an ihre Grenzen. In CEFMON wird ein numerische Optimierungsverfahren eingesetzt, dass zu einem Korrelationskoeffizienten von ca. 0.65 führt. Den Autoren ist kein besseres Resultat zu diesem Problem bekannt. Das Ergebnis einer (immer noch verbesserungsfähigen) Suboptimierung ist in den folgenden Abbildungen zu sehen.

Die Abb. 7-1 zeigt im oberen Bild die Funktion *fhm*, darunter die Gewichtsfunktion, mit der *fhm* multipliziert wird und darunter das Ergebnis, die Funktion FRQ, wie sie in der WAS-Index-Formel benutzt wird. Analoges gilt für die Abb. 7-2 und 7-3.

Die Tatsache, dass auch negative Gewichtungen vorkommen, stört nicht. Dadurch wird im Optimierungsverfahren erreicht, dass der WAS-Score an kritischen Stellen entsprechend kleine Werte annimmt. Bei der Betrachtung der Kurven sollte man beachten, dass ein Optimierungsprogramm nur in Variablenregionen zur Wirkung kommt, die auch einen Einfluss auf den Korrelationskoeffizienten haben. Anders ausgedrückt: einige „Zacken“ der Kurven dürften ohne Bedeutung sein. Es sei an Tabelle 3 erinnert, die zeigt, wo die mittleren Werte der Variablen liegen.

**Abb. 7-1**

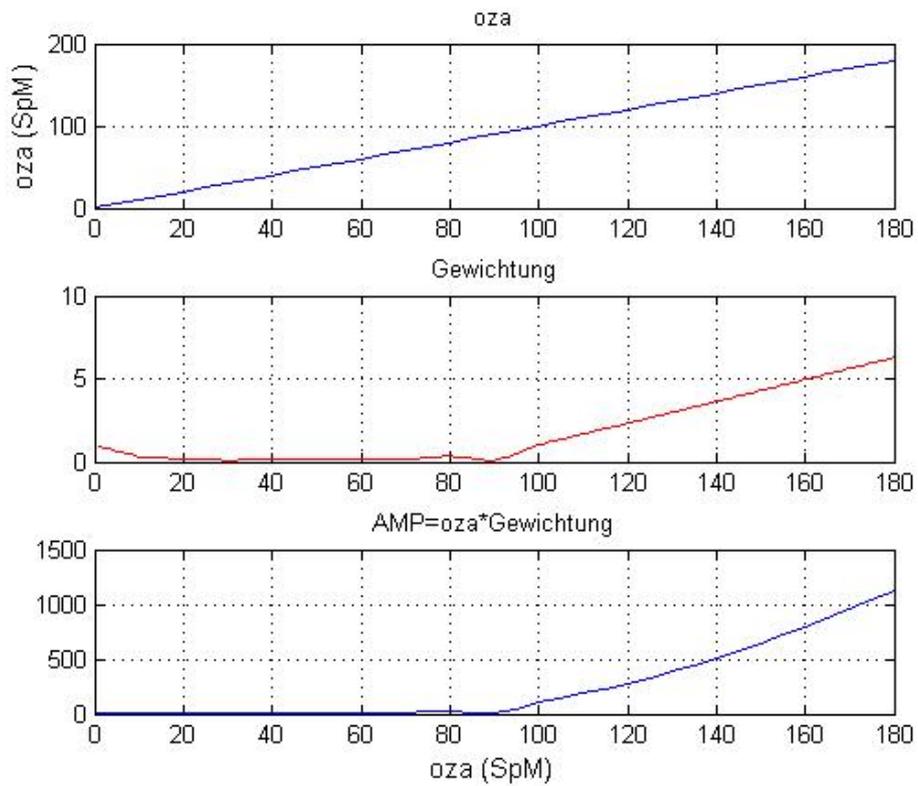


Abb. 7-2

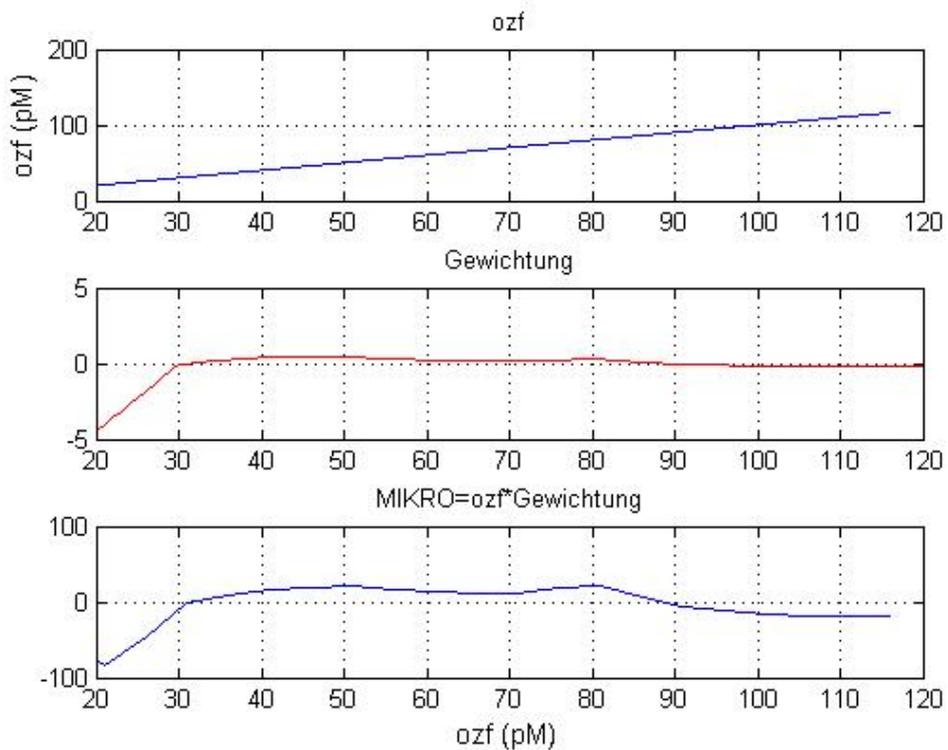


Abb. 7-3

Abb. 7-4 zeigt die Verteilung der WAS-Score-Werte, normiert auf Werte zwischen 0 und 1, von 478 Fällen des Archivs. Dazu ist zum Vergleich die Dichtefunktion der Normalverteilung (mit gleichem Mittelwert und Streuung) zu sehen.

Rechnet man nun noch die WAS-Score-Werte durch eine weitere lineare Transformation in pH-Werte um, so kann man ein Punktediagramm von prognostizierten pH-Werten (X-Achse) und gemessenen pH-Werten (Y-Achse) erzeugen.

Das Resultat ist in Abb. 7-5 zu sehen. Die Umrechnung erfolgt nach derselben Methode wie beim Q-Score. Die Ausgleichsgerade nach Gauß stimmt hier, wie man nachrechnen kann, mit der Diagonalen überein.

Abb. 7-6 zeigt die Verteilung der Archiv-Fälle bezüglich den gemessenen pH-Werten (pHNA) und den mit dem WAS-Score berechneten pH-Werten. Optisch ist kein Unterschied zu erkennen.

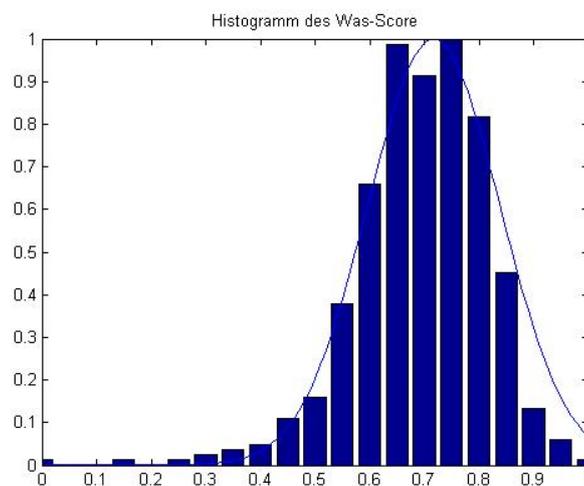


Abb. 7-4: Relative Häufigkeit des WAS-Scores (normiert) bei 478 Fällen

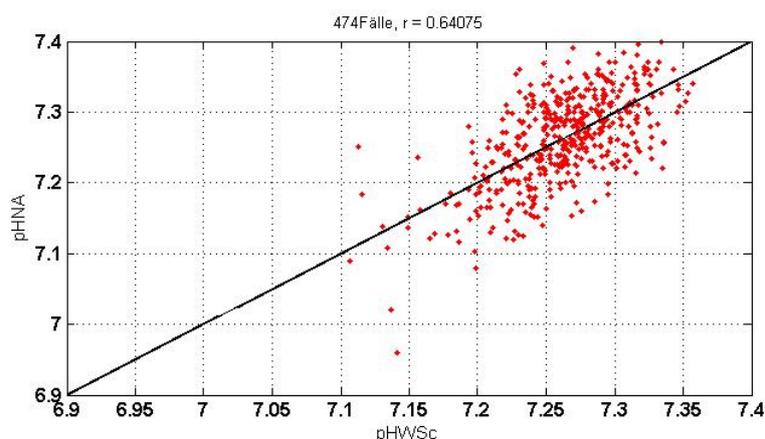


Abb. 7-5

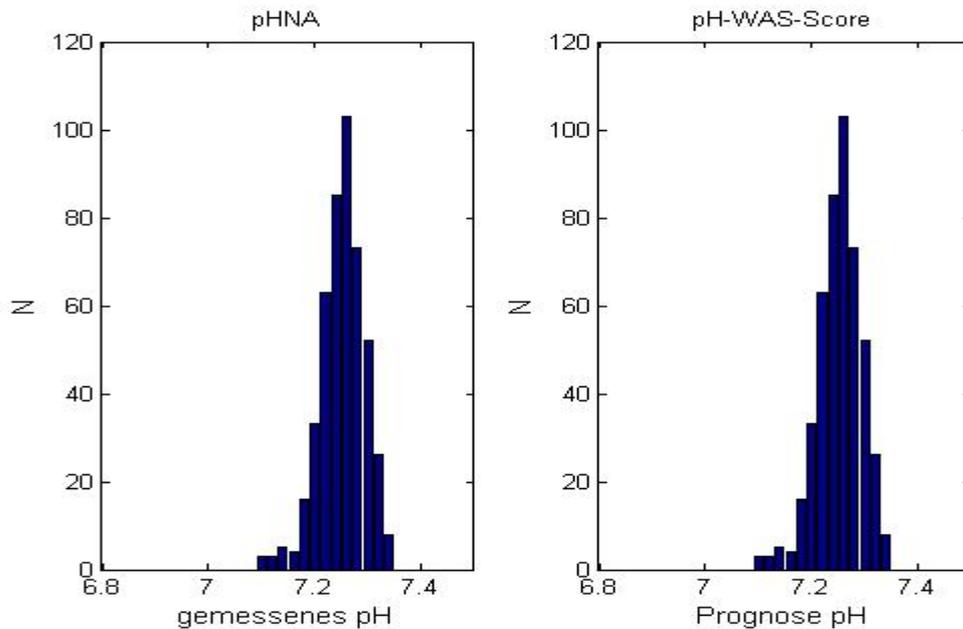


Abb. 7-6

Test auf Azidose

Eine grundlegende Frage der CTG-Analyse ist, ob man anhand eines 30-Minuten-CTGs eine Azidose zuverlässig erkennen kann. Da unser Datenarchiv die Messwerte des pH-Werts im Augenblick der Geburt und die FHF-Kurve der letzten 30 Minuten enthält, kann man den WAS-Score als Testverfahren zum Test auf Azidose benutzen. Diesen für den Augenblick der Geburt verifizierten Test kann man dann anwenden sobald man die FHF-Daten von 30 Minuten hat. Wir gehen dabei davon aus, dass die biochemischen Vorgänge stets nach gleichen Gesetzmäßigkeiten ablaufen.

Wir legen eine pH-Wert-Schranke fest, die Azidosen von Nichtazidosen trennt. Der Test ist positiv (p), wenn er eine Azidose anzeigt, sonst negativ (\bar{p}). Da der Sachverhalt Azidose (a) oder Nichtazidose (\bar{a}) vorliegen kann, auf den der Test positiv oder negativ reagieren kann, ergeben sich vier mögliche Kombinationen beim Vorliegen eines Testresultats:

Realität/Test	positiv (p)	negativ (\bar{p})
keine Azidose (\bar{a})	FP	TN
Azidose (a)	TP	FN

Tabelle 4

TP steht für **true positive** und bezeichnet die Tatsache dass das positive Ergebnis des Tests wahr ist. **FN** steht für **false negative** und bezeichnet die Tatsache, dass das negative Ergebnis des Tests falsch ist. Analog verhält es sich mit den anderen Bezeichnungen. Bei einem Test liegt also stets genau einer der vier Fälle vor. Liegen viele Testergebnisse vor, kann man die vier Felder mit den relativen Häufigkeiten der Testreihe füllen. Mit **tpr** bezeichnen wir dann die relative Anzahl der TP-Werte der Testmenge, analog **fpr**, **fnr** und **tnr** (siehe unten).

In unserem Fall führen wir folgenden Test durch:

Wir legen eine (klinisch gewollte) pH-Wert-Schranke pH_0 fest, die Azidosen von Nichtazidosen trennt. Dann legen wir eine zweite Schranke W_0 für den Was-Index fest, die ein positives Testergebnis (kleiner Was-Index) von einem negativen Testergebnis (großer Was-Index) trennt. Wir sagen:

1. Der Test ist TP, wenn der gemessene pH-Wert kleiner als pH_0 ist und der WAS-Score kleiner als W_0 ist.
2. Der Test ist FN, wenn der gemessene pH-Wert kleiner als pH_0 ist und der WAS-Score größer als W_0 ist.
3. Der Test ist FP, wenn der gemessene pH-Wert größer als pH_0 ist und der WAS-Score kleiner als W_0 ist.
4. Der Test ist TN, wenn der gemessene pH-Wert größer als pH_0 ist und der WAS-Score größer als W_0 ist.

In Abb. 8-1 sind alle Fälle des Archivs grafisch eingetragen, dazu die Geraden, die den Werten von $pH_0 = 7.2$ und $W_0 = 0.9$ (als Beispiel) entsprechen. Dadurch wird die Grafik in vier Quadranten aufgeteilt, die entsprechend der Tabelle 4 bezeichnet sind.

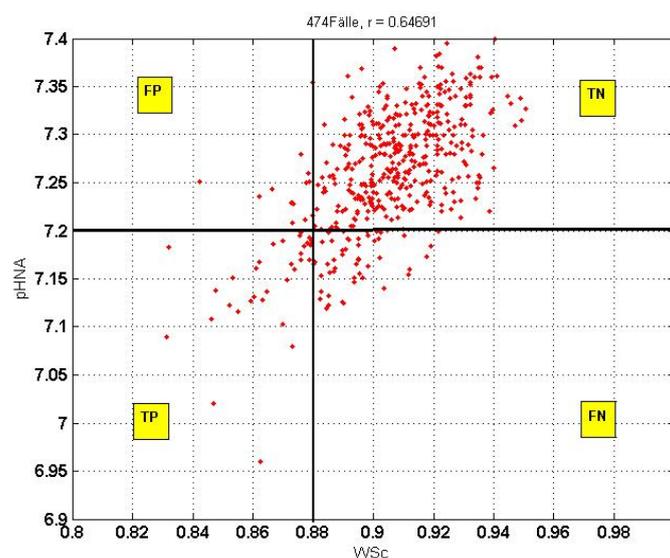


Abb. 8-1

Wir berechnen die relativen Häufigkeiten:

$$tpr = TP / (TP + FN)$$

$$fnr = FN / (TP + FN)$$

$$fpr = FP / (FP + TN)$$

$$tnr = TN / (FP + TN)$$

tpr wird auch als **Sensitivität** bezeichnet, tnr als **Spezifität**.

Wir gehen zu den in der Wahrscheinlichkeitsrechnung üblichen Bezeichnungen über.

Statt tpr schreiben wir $P(p|a)$. Damit wird die **bedingte** Wahrscheinlichkeit bezeichnet, dass der Test positiv ist **unter der Bedingung**, dass eine Azidose vorliegt.

Statt tnr schreiben wir $P(\bar{p}|\bar{a})$. Damit wird die bedingte Wahrscheinlichkeit bezeichnet, dass der Test negativ ist unter der Bedingung, dass keine Azidose vorliegt.

Statt fpr schreiben wir $P(p|\bar{a})$. Damit wird die bedingte Wahrscheinlichkeit bezeichnet, dass der Test positiv ist unter der Bedingung, dass keine Azidose vorliegt.

Statt fnr schreiben wir $P(\bar{p}|a)$. Damit wird die bedingte Wahrscheinlichkeit bezeichnet, dass der Test negativ ist unter der Bedingung, dass eine Azidose vorliegt.

Dabei ist nach der Bayesschen Formel $P(p|a) = \frac{P(a \cap p)}{P(a)}$, usw.

Wir haben zu unterscheiden zwischen $P(p|a)$ und $P(a|p)$. $P(p|a)$ ist die Wahrscheinlichkeit dafür, dass der Test eine Azidose erkennt. $P(a|p)$ ist die Wahrscheinlichkeit dafür, dass bei positivem Test tatsächlich eine Azidose vorliegt. Dieser Wert interessiert uns und sollte möglichst hoch sein.

Da in der Bayesschen Formel a und p vertauschbar sind, gilt auch

$$P(a|p) = \frac{P(a \cap p)}{P(p)} \text{ und somit}$$

$$P(a|p) = \frac{P(a \cap p)}{P(p)} = \frac{P(a \cap p)}{P(a)} \cdot \frac{P(a)}{P(p)} = P(p|a) \cdot \frac{P(a)}{P(p)}$$

Wir betrachten ein numerisches Beispiel, das die Nützlichkeit dieser Formel und des Tests unterstreichen soll.

Wir wählen $pH_0 = 7.2$ und $W_0 = 0.88$. Unser Datenmaterial ergibt, dass dann

Realität/Test	positiv (p)	negativ (\bar{p})
keine Azidose (\bar{a})	14	382
Azidose (a)	32	46

Tabelle 5

Und somit $\text{tpr} = P(p|a) = 0.410$ und $\text{fpr} = P(p|\bar{a}) = 0.035$

ist. Ferner sei $P(a) = 0.18$ die Wahrscheinlichkeit, dass (bei zufälliger Auswahl) eine Azidose vorliegt. $P(p)$ berechnen wir wie folgt:

$$P(p) = P(p|a) \cdot P(a) + P(p|\bar{a}) \cdot P(\bar{a}) = 0.419 \cdot 0.18 + 0.035 \cdot 0.72 = 0.1028$$

Damit ergibt sich $P(a|p) = 0.7181$.

Das bedeutet, dass gegenüber dem zufälligen Antreffen einer Azidose mit 18% Wahrscheinlichkeit nach Anwendung des Tests eine Azidose mit einer Wahrscheinlichkeit von 72,5% erkannt wird.

ROC-Analyse

Die ROC-Analyse (ROC = Receiver-Operating-Characteristics) wurde während des 2. Weltkriegs zur Rauschunterdrückung in der Radartechnik entwickelt. Sie dient dazu, das Verhalten von Diagnose-Systemen zu analysieren. Das Verfahren wird seit den 1960er Jahren auch in der Medizin als Analyse-Verfahren in der Krankheitsdiagnostik eingesetzt.

In Abb. 8-1 beginnen die Azidosen bei einem pH-Wert von 7.2. Diesen Wert auf der y-Achse wollen wir den **Schwellenwert** des Verfahrens nennen. Wenn man die vertikale Gerade, in Abb. 8-1 von links nach rechts verschiebt, bekommt man eine Schar von Werten für die Spezifität und Sensitivität, die man gegeneinander auftragen kann. Die ist in Abb. 9-1 geschehen.

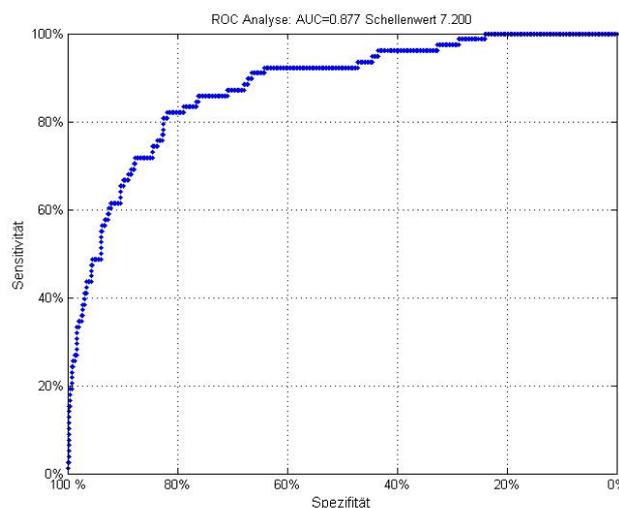


Abb. 9-1

Die Kurve heißt ROC-Kurve. Die Fläche unter der Kurve AUC (area under curve) ist ein Maß für die Qualität des Testverfahrens. Bei einem 100% sicheren Testverfahren ist die ROC-Kurve identisch mit der linken vertikalen Achse und der oberen horizontalen Achse. Die Fläche unter der ROC-Kurve ist dann 1, d.h. die Gesamtfläche des Quadrats.

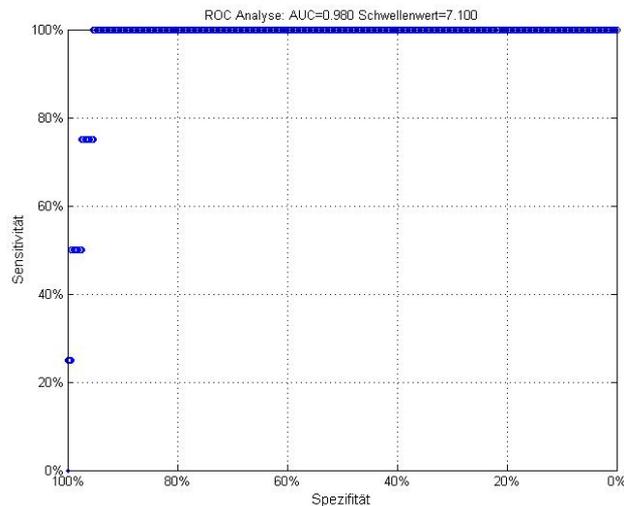


Abb. 9-2

Bei einem Schwellenwert von 7.1 ergibt sich die ROC-Kurve in Abb. 9-2. AUC ist auf 0.980 gestiegen.

Gesucht ist natürlich der „richtige“ Schwellenwert für den Beginn einer Azidose. Dabei ist auf die Falsch-Negativ-Rate fnr zu achten, die zumindest für die Daten des Archivs Null sein sollte. $fnr=0$ ist gleichbedeutend mit einer Sensitivität von 100%, die in Abb. 9-1 mit einer Spezifität von ca. 25% erreicht wird, in Abb. 9-2 aber mit ca. 95%.

Aus Abb. 8-1 sieht man, welche Möglichkeiten es gibt das Achsenkreuz zu legen, damit der FN-Quadrant keinen Punkt enthält, fnr also null ist.

Abb. 9-3 unterscheidet sich von Abb. 8-1 nur dadurch, dass die Werte auf der x-Achse in pH-Werte umgerechnet sind. Dies ist eine für unsere Überlegungen unschädliche Umrechnung, die uns aber ermöglicht, die gemessenen pH-Werte und die durch den WAS-Score prognostizierten pH-Werte zu vergleichen. Wäre die Prognose zu 100% richtig, würden alle Punkte in Abb. 8-1 und Abb. 9-3 auf einer Geraden liegen. Die Tatsache, dass sie nicht alle auf einer Geraden liegen, wird durch den Korrelationskoeffizienten beider Werte gemessen, der bei uns bei 0.647 liegt.

Man erkennt aus Abb. 9-3, dass man bei einem Azidose-Schwellenwert von 7.1 (grafisch darzustellen als eine Gerade parallel zur x-Achse durch den y-Wert 7.1) und einem Prognosewert von ca. 7.19 (grafisch darzustellen als eine Gerade parallel zur y-Achse durch den x-Wert 7.19) einen leeren FN-Quadranten hat, d.h. in dem FN-Quadranten liegt dann keiner der Punkte. Bei einem Azidose-Schwellenwert von 7.2 (y-Wert) liegt die Prognoseschwelle bei ca. 7.3 (x-Wert).

Man erkennt: bei einem exakten Prognoseverfahren wären die beiden Werte gleich und die beiden Geraden würden die vier Quadranten wie in Abb.9-3

erzeugen. Aber in Abb.9-3 ist der FN-Quadrant nicht leer. Die Prognose würde einige Fälle aus dem Archiv nicht als Azidose erkennen.

Ideal ist ein Schwellenwert mit dazu passendem Prognosewert, für die der FN-Quadrant leer ist **und** der FP-Quadrant möglichst wenige Punkte enthält. Ohne zu rechnen sieht man aus der Grafik, dass das bei einem Schwellenwert von ungefähr 7.12 (y-Wert) und einem Prognosewert von ca. 7.19 (x-Wert) der Fall ist. Ein „Alarm“ unter der Geburt sollte also ausgelöst werden, wenn der mit Hilfe des WAS-Score vorhergesagte pH-Wert unter 7.19 fällt.

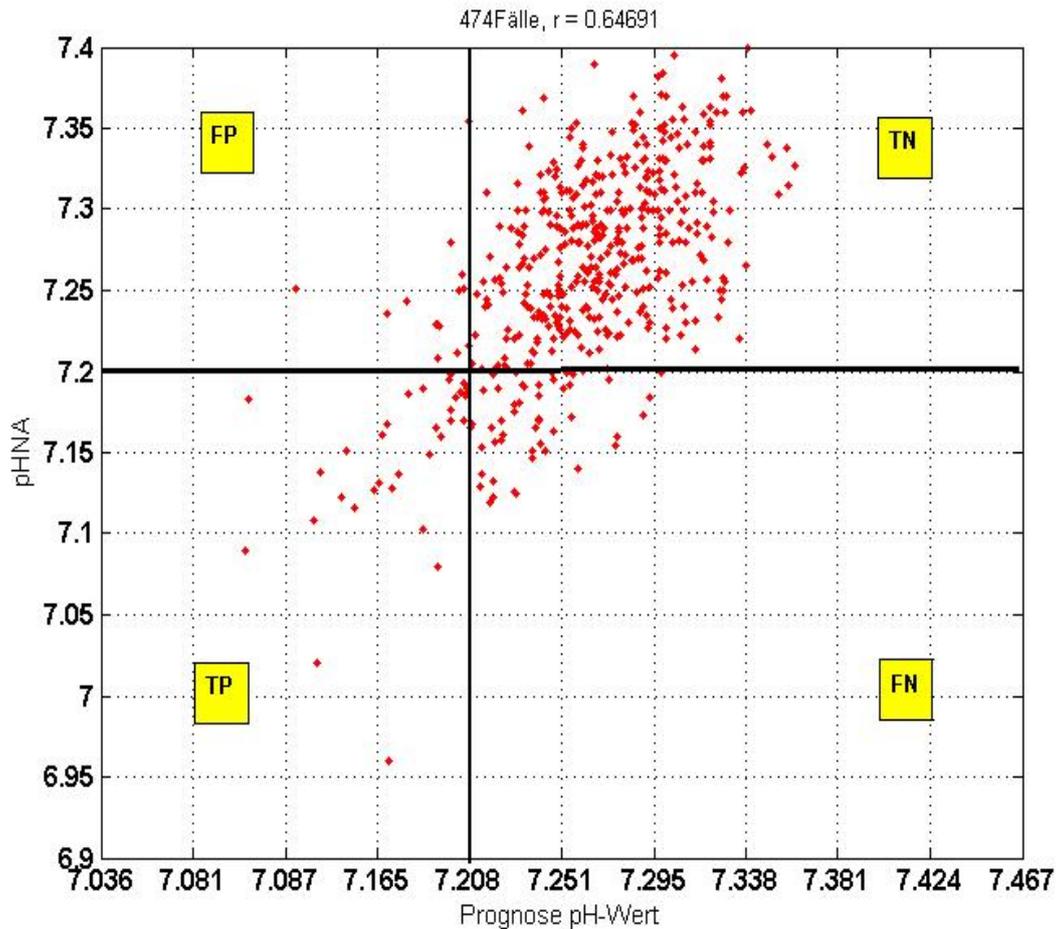


Abb. 9-3

Informationen

Veröffentlichungen mit Bezug auf CEFMON

Roemer V.M., Walden, R. Quantitative Kardiotokographie. Geburtshilfe und Frauenkunde 64 2004, 805-815

Roemer V.M., Walden, R. Neues Reparaturprogramm für FHF-Kurven. Frauenarzt 48,2007, Nr.10, 982-989

Roemer V.M., Walden, R. Neue Wege der Kardiotokografie. Gyn 9, 2004, 430-437

Roemer V.M., Walden, R. Quantitative Kardiotokographie – wie sieht sie aus und was dürfen wir erwarten? Z Geburtsh Neonatol 2006; 210: 77-91

Roemer V.M. CTG: Mikrofluktuation. Z Geburtsh Neonatol 2004; 208: 210-219

Roemer V.M. Computergestützte CTG-Bewertung mit Berücksichtigung der mütterlichen Atmung Geburtshilfe und Frauenheilkunde 63 (2003) 766-773.

Roemer V. M.: Quantitative CTG-Bewertung sub partu mit einem neuen CTG-Score: Wie gut sind die Korrelationen mit den Parametern des fetalen Säure-Basen-Haushalts im Nabelschnurblut?. Z Geburtsh Neonatol 2003; 207: 121-126.

Roemer V. M.: Die fetale Überwachung - ein Plädoyer für die Kardiotokographie. Geburtshilfe und Frauenheilkunde 58 (1998) 1-10.

Roemer V. M.: Der Kardiotokographiescore nach Hammacher in der elektronischen Analyse. Gynäkol Geburtshilfliche Rundsch 2000;40:20-28

Roemer V. M.: Computerassistierte Analyse von intrapartalen Kardiotokogrammen. Zentralblatt Gynökol 124 (2002) 317-325.

Walden, R. Entwicklung eines computergestützten Verfahrens zur Analyse von fetalen Herzfrequenzsignalen.

Roemer V. M., Walden, R: Sensitivity, Spezifity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios for electronic foetal heart rate monitoring using new evaluation techniques.

Anschriften

Fachhochschule Bielefeld
Fachbereich Ingenieurwissenschaften und Mathematik
FSP AMMO
Am Stadtholz 24
33609 Bielefeld
Tel.: +49.521.106-7403
Fax: +49.521.106-7176
E-Mail: ammo@fh-bielefeld.de
Internet: www.fh-bielefeld.de/ammo

Prof. Dr. Volker Michael Roemer **Institut für feto-maternale Medizin**

Benekestrasse 2
D-32756 Detmold
Tel.: 0049(0)5231 35842
E-mail: VMR.DR@t-online.de

Prof. Dr. Rainer Walden

Hubertusstr. 21
D-33129 Delbrück
Tel.: 0049(0)5250 7519
E-mail: rainer.walden@web.de

ISSN 2198-4824

Herausgeber: Sprecher FSP AMMO
Fachhochschule Bielefeld