

**Planung, Durchführung und Auswertung
einer klinischen Studie zur Bewertung und
zum Vergleich von Alarmsystemen
in der Intensivmedizin**

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik der Technischen Universität Dortmund
vorgelegt von

Silvia Kuhls

Dortmund 2008

1. Gutachter: Prof. Dr. Ursula Gather

2. Gutachter: Prof. Dr. Roland Fried

Tag der mündlichen Prüfung: 30. Juni 2008

Inhaltsverzeichnis

1	Einleitung	1
2	Intensivmedizinisches Online-Monitoring	3
2.1	Klinische Praxis	3
2.2	Robuste Signalextraktion	4
2.3	Weitere Ansätze im Überblick	6
3	Planung der klinischen Studie	8
3.1	Studienziele	9
3.2	Patientenkollektiv	10
3.3	Überblick über das Datenerhebungskonzept	11
3.4	Planung der statistischen Analyse	13
4	Datenerhebung vom Monitorsystem	16
4.1	Das Infinity Monitoring System [®]	16
4.1.1	Alarmer	17
4.1.2	Häufig überwachte Vitalparameter	19
4.2	Die Datenerfassung	20
4.3	Datenerhebungskonzepte anderer Studien	21
5	Klinische Annotationen	23
5.1	Hilfsmittel	23
5.1.1	Videoaufzeichnung	23
5.1.2	Eingabemaske	24
5.1.3	Echtzeitkurven	26
5.2	Bewertung von Alarmsituationen	27
5.3	Annotationskonzepte anderer Studien	30
6	Auswertungsmethoden	35
6.1	Methoden für die Bewertung und den Vergleich von Diagnoseverfahren	36

6.1.1	Simultaner Vergleich von Sensitivitäten und Spezifitäten	39
6.1.2	Berücksichtigung von Abhängigkeiten	43
6.2	Übertragung der Methodik auf den Vergleich von Alarmsystemen	49
6.2.1	Bestimmung der Sensitivität	49
6.2.2	Bestimmung der Spezifität	51
6.2.3	Auswahl von Validierungsmethodiken	55
6.3	Weitere Beurteilungskriterien	57
7	Der Referenzdatensatz	61
7.1	Beobachtungsdauer und Alarmhäufigkeit	61
7.2	Vitalparameter	63
7.3	Alarmmeldungen	64
7.4	Annotationen	66
8	Regeln für die Offline-Anwendung neuer Alarm-Algorithmen	72
8.1	Alarmreihenfolge	73
8.2	Alarm-Deaktivierungszeiträume	73
9	Vergleich von Alarm-Algorithmen	77
9.1	Auswahl neuer Alarm-Algorithmen	77
9.1.1	Arterielle Blutdrücke	79
9.1.2	Herzfrequenz	95
9.1.3	Sauerstoffsättigung	99
9.2	Validierung neuer Alarm-Algorithmen	103
9.2.1	Arterielle Blutdrücke	104
9.2.2	Herzfrequenz	106
9.2.3	Sauerstoffsättigung	108
9.3	Diskussion der Ergebnisse	110
10	Zusammenfassung und Ausblick	113
	Literaturverzeichnis	116
A	Studienprotokoll	120
B	Ergänzende Deskription	145
C	Ergänzungen zum Alarmsystem-Vergleich	163

Kapitel 1

Einleitung

Die Patientenüberwachung auf medizinischen Intensivstationen erfolgt mittels moderner Monitoring-Geräte. Diese messen und erfassen die Werte hämodynamischer Vitalparameter und alarmieren das medizinische Personal, wenn die Werte auf besorgniserregende Änderungen des Patientenzustandes oder technische Probleme hindeuten. Verschiedene Studien zeigen, dass derzeitig verwendete Alarmsysteme kritische Situationen zwar zuverlässig erkennen, jedoch gleichzeitig sehr viele Fehlalarme produzieren (Koski et al., 1990, Lawless, 1994, Tsien und Fackler, 1997, Chambrin et al., 1999). Die Folge ist eine Desensibilisierung des medizinischen Personals und somit die Gefahr, dass auf wichtige Alarme nicht rechtzeitig reagiert wird.

Im Rahmen des Projekts „Zeitreihenanalytische Methoden zur Behandlung von Online-Monitoring-Daten aus der Intensivmedizin“ im Sonderforschungsbereich 475 „Komplexitätsreduktion in multivariaten Datenstrukturen“ an der Technischen Universität Dortmund werden statistische Methoden zur Verbesserung der bestehenden Alarmsysteme erforscht. Insbesondere Alarm-Algorithmen basierend auf Verfahren der robusten Signalextraktion bieten das Potenzial, die Anzahl der überflüssigen Fehlalarme zu reduzieren, ohne kritische Situationen zu verpassen. Vor einer Anwendung neuer Alarm-Algorithmen in der Praxis ist deren Eignung und Unbedenklichkeit adäquat zu überprüfen. Insbesondere ist ein Vergleich mit dem zurzeit verwendeten Alarmsystem vorzunehmen. Dies geschieht derzeit im Rahmen einer klinischen Studie, deren Planung, Durchführung und Auswertung Gegenstand der vorliegenden Arbeit ist.

Das folgende Kapitel gibt zunächst eine Einführung in die klinische Praxis des intensivmedizinischen Online-Monitorings und erläutert das Prinzip der dort standardmäßig verwendeten Schwellwertalarme. Anschließend werden Methoden der robusten Signal-

extraktion vorgestellt, welche als Basis für eine Alternative zur einfachen Schwellwertalarmgebung dienen sollen. Es folgt eine Übersicht über weitere aus der Literatur bekannte Ansätze zur Verbesserung derzeitig verwendeter Alarmsysteme.

Im dritten Kapitel wird die Planung der klinischen Studie beschrieben. Diese umfasst die Festlegung der genauen Studienziele, die Auswahl des Patientenkollektivs sowie die Erstellung eines Datenerhebungskonzeptes und eines statistischen Analyseplans. Im Rahmen der Studie wird in Zusammenarbeit mit dem Universitätsklinikum Regensburg ein klinisch annotierter Referenzdatensatz erstellt, auf dessen Basis neue Alarm-Algorithmen offline validiert werden können. Hierfür werden Daten vom Monitorsystem benötigt. Eine Beschreibung des am Studienort Regensburg verwendeten Infinity Monitoring Systems[®] (Dräger Medical, Lübeck) und detaillierte Informationen zur Datenerhebung finden sich in Kapitel 4.

Auf Basis von Videoaufzeichnungen und grafisch aufbereiteten Monitoringdaten erfolgt eine Bewertung der während der Datenaufzeichnung aufgetretenen Alarmsituationen. Insbesondere wird vermerkt, ob ein sogenannter „alarmrelevanter Zustand“ vorliegt. Dies ist für die spätere Bewertung neuer Alarm-Algorithmen von Bedeutung, da diese daran gemessen werden sollen, wie sicher sie derartige alarmrelevante Zustände erkennen. Eine ausführliche Beschreibung der Annotationsweise sowie eine Übersicht über Annotationskonzepte anderer Studien findet sich in Kapitel 5.

Kapitel 6 beschäftigt sich mit den statistischen Auswertungsmethoden. Als maßgebliche Beurteilungskriterien für die zu vergleichenden Alarm-Algorithmen werden die Sensitivität in Bezug auf die Erkennung alarmrelevanter Zustände und die Spezifität betrachtet. Für den simultanen Vergleich von Sensitivitäten und Spezifitäten werden Analysemethoden für die Bewertung und den Vergleich diagnostischer Verfahren benötigt, welche in Abschnitt 6.1 erläutert werden. In Abschnitt 6.2 erfolgt die notwendige und neue Übertragung der Methodik auf den Alarmsystem-Kontext. Weitere Beurteilungskriterien für Alarmsysteme werden in Abschnitt 6.3 vorgestellt.

In Kapitel 7 findet sich eine deskriptive Analyse des Referenzdatensatzes hinsichtlich Beobachtungsdauer und Alarmhäufigkeit, überwachten Vitalparametern, aufgetretenen Alarmmeldungen und Annotationen. Auf Basis des Referenzdatensatzes werden offline neue Alarme generiert. Allgemeine Regeln hierfür werden in Kapitel 8 beschrieben.

Der Vergleich neuer und alter Alarm-Algorithmen erfolgt in Kapitel 9. Zunächst werden anhand einer Teilstichprobe des Referenzdatensatzes die genauen Einstellungen und Regeln der neuen Alarm-Algorithmen bestimmt. Die ausgewählten Verfahren werden anschließend anhand der übrigen Daten des Referenzdatensatzes validiert.

Kapitel 10 fasst die Ergebnisse der Arbeit zusammen.

Kapitel 2

Intensivmedizinisches Online-Monitoring

2.1 Klinische Praxis

Alarmsysteme für das hämodynamische Monitoring von Intensivpatienten können eine Vielzahl unterschiedlicher Alarme auslösen. Neben Alarmen, die auf rein technische Probleme während der Überwachung aufmerksam machen und Alarmen, die spezielle Arrhythmie-Ereignisse durch eine Analyse des Elektrokardiogramms feststellen, sind dies sogenannte Schwellwertalarme. Von allen Alarmarten treten diese in der Praxis am häufigsten auf.

Schwellwertalarmsysteme bieten dem Mediziner die Möglichkeit für jeden Patienten individuell und für jeden interessierenden Vitalparameter einen Bereich festzulegen, in dem die gemessenen Werte als unbedenklich eingestuft werden. Bei Abweichungen von diesem „Normbereich“ werden vom Monitoringsystem Alarme ausgelöst. Obwohl auch innerhalb der eingestellten Alarmgrenzen möglicherweise Ereignisse auftreten können, die medizinisch von Interesse sind – etwa plötzliche Niveauänderungen oder langsame Trends – hat sich das Prinzip der auf festen Grenzen basierenden Alarme durchgesetzt und ist in jedem kommerziell verfügbaren Monitoring-Gerät implementiert.

Schwellwertalarme basieren auf den gemessenen Zeitreihenwerten der Vitalparameter. In solchen Zeitreihen treten häufig Messartefakte und kurzfristige Schwankungen auf, welche die Ursache für viele klinisch unbedeutende Schwellwertalarme sind. Eine Möglichkeit diese Art von Alarmen zu vermeiden besteht in einer besseren Vorverarbeitung der Daten, etwa durch eine geeignete Filtermethode. Anstelle der Original-

Messwerte können dann die geglätteten und von Rauschen und Ausreißern bereinigten Daten mit den Alarmgrenzen verglichen werden.

Filterverfahren, wie z.B. das gleitende Mittel oder der gleitende Median sind allerdings entweder nicht robust gegenüber Ausreißern oder sie liefern bei Zeitreihen mit Trend kein glattes gefiltertes Signal. Eine alternative Möglichkeit der Datenvorverarbeitung bietet eine robuste Online-Signalextraktion, die im folgenden Abschnitt vorgestellt wird.

2.2 Robuste Signalextraktion

Bei der robusten Signalextraktion erfolgt die Schätzung des der Zeitreihe zugrunde liegenden Signals mittels einer robusten Regressionsmethode, welche in einem gleitenden Zeitfenster angewandt wird (Davies, Fried und Gather, 2004).

Für die Zeitreihe $(y_t)_{t \in \mathbb{N}}$ wird als einfaches Arbeitsmodell angenommen, dass sie sich additiv aus einem glatten Signal μ_t , sowie Rauschen und Ausreißern zusammensetzt. Weiter wird unterstellt, dass sich die Werte in einem Fenster $(y_{t-m}, \dots, y_t, \dots, y_{t+m})$ der Länge $2m + 1$ durch einen linearen Trend approximieren lassen:

$$y_{t+i} = \mu_t + \beta_t i + \varepsilon_{t,i}, \quad i = -m, \dots, m.$$

Der Wert der durch eine geeignete Regressionsmethode angepassten Gerade zum mittleren Zeitpunkt des Fensters $\hat{\mu}_t$ ist eine Möglichkeit zur Schätzung des Signals, allerdings mit einer Verzögerung von m Zeitpunkten. Eine „Online“-Schätzung erhält man durch Verwendung des Wertes der Regressionsgerade am Ende des Fensters

$$\hat{\mu}_{t+m}^{online} = \hat{\mu}_t + \hat{\beta}_t m$$

(Gather, Schettlinger und Fried, 2006). Eine grafische Veranschaulichung des Prinzips findet sich in Abbildung 2.1.

In einem Vergleich verschiedener robuster Regressionsverfahren hinsichtlich Bias, Varianz, MSE, Bruchpunkt und Rechenzeit schneidet der von Siegel (1982) vorgeschlagene Repeated Median

$$\hat{\beta}_t^{RM} = \operatorname{med}_{i=-m}^m \left\{ \operatorname{med}_{j \neq i} \frac{y_{t+i} - y_{t+j}}{i - j} \right\}$$

$$\hat{\mu}_t^{RM} = \operatorname{med}_{i=-m}^m \left\{ y_{t+i} - \hat{\beta}_t^{RM} i \right\}$$

insgesamt am besten ab (Davies, Fried und Gather, 2004, Gather, Schettlinger und Fried, 2006). Die Berechnung von $\hat{\beta}_t^{RM}$ lässt sich anschaulich wie folgt erklären: Für

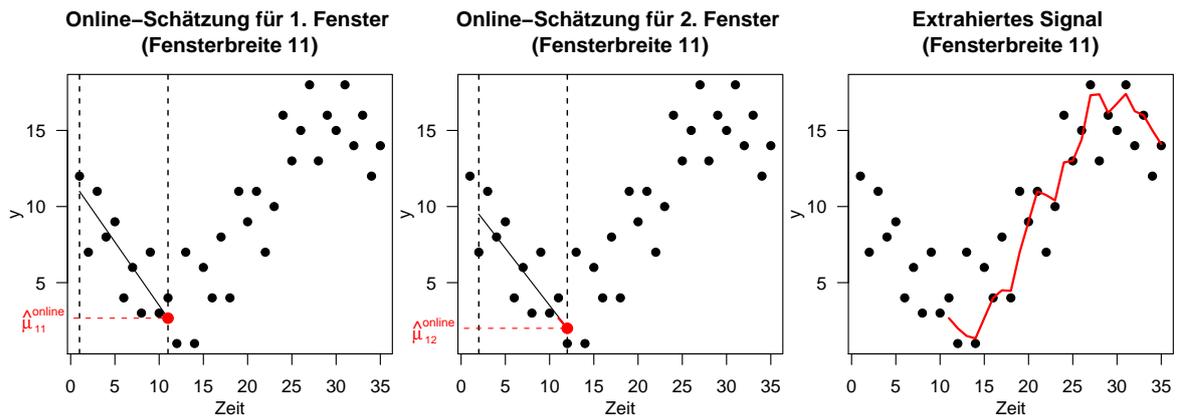


Abbildung 2.1: Veranschaulichung des Prinzips der Online-Signalextraktion

jeden Zeitpunkt $t + i$ im Fenster wird der Median der Steigungen aller Geraden bestimmt, die entstehen, wenn der Punkt $(t + i, y_{t+i})$ mit den anderen Punkten im Fenster verbunden wird. $\hat{\beta}_t^{RM}$ ist dann wiederum der Median dieser $2m + 1$ medianen Steigungen.

Modifikationen der einfachen Signalextraktion durch zusätzliche Regeln zur Ausreißerersetzung und Shift-Erkennung werden in Fried (2004) untersucht. Eine Signalextraktionsprozedur mit adaptiver Fensterbreite wird in Gather und Fried (2004) vorgestellt und von Schettlinger, Fried und Gather (2008) weiterentwickelt. Hier wird keine feste Fensterbreite vorgegeben, sondern diese kann zwischen einer minimalen und einer maximalen Größe variieren und wird online und abhängig von der Datensituation automatisch gewählt.

In der vorliegenden Arbeit werden auf Basis klinisch annotierter Daten offline signalextraktionsbasierte Schwellwertalarme generiert und mit den einfachen Schwellwertalarmen des derzeitigen Alarmsystems verglichen. Im Rahmen der Untersuchung ist zu klären, für welche Vitalparameter Signalextraktions-Alarme geeignet sind, wie die Fensterbreite jeweils zu wählen ist und ob Zusatzregeln für die Alarmgebung nötig sind. Die Güte des Abschneidens eines ausgewählten neuen Alarm-Algorithmus ist anschließend danach zu beurteilen, wieviele der klinisch bedeutsamen Situationen vom neuen Verfahren erkannt und wie viele unbedeutende Alarme gleichzeitig durch die Filterung vermieden werden.

2.3 Weitere Ansätze im Überblick

In der Literatur finden sich viele Vorschläge zur Verbesserung der Alarmgebung und Entscheidungsfindung im intensivmedizinischen Online-Monitoring. Die meisten Arbeiten kommen aus der Statistik und der Künstlichen Intelligenz. Das Spektrum der Verfahren umfasst Signalvorverarbeitung, Prozesskontrolle und Mustererkennung, aber auch Maschinelles Lernen, Neuronale Netze oder Fuzzy Logic. Eine Übersicht über die wichtigsten Ansätze und ihr Abschneiden bei der Anwendung auf intensivmedizinische Daten findet sich in Imhoff und Kuhls (2006).

Zwei Vorschläge beruhen ebenfalls auf der Idee, für die Schwellwertalarmgebung nicht die Originalbeobachtungen, sondern ein daraus gewonnenes Signal zu verwenden. Mäki-virta et al. (1991) überwachen intensivmedizinische Zeitreihen mittels zweier mediangefilterter Signale. Dabei dient der erste Median-Filter mit einer kurzen Fensterbreite von 31 Sekunden in Kombination mit sehr weiten Alarmgrenzen zur Entdeckung plötzlicher und extremer Signalveränderungen. Der zweite Median-Filter operiert mit einer größeren Fensterbreite (ab zwei Minuten) und engeren Alarmgrenzen, um Alarme aufgrund kurzfristiger Fluktuationen zu vermeiden. Eine Anwendung auf Daten von zehn Patienten, gemessen im 5-Sekundentakt, mit Beobachtungsdauern zwischen 3 und 5.5 Stunden ergibt eine Steigerung des Anteils richtiger Alarme von 12% auf 49% im Vergleich zu einfachen Schwellwertalarmen. Die Häufigkeit falscher Alarme reduziert sich von 12 auf im Durchschnitt 4.5 Alarme pro Stunde, ohne dass wichtige klinische Situationen verpasst werden.

Charbonnier, Becq und Biot (2004) zerlegen die Originalzeitreihe online in lineare Trends unterschiedlicher Dauer. Diese sogenannte „Online-Segmentierung“ der Zeitreihe wird erreicht durch Fortführung eines mittels KQ-Regression angepassten linearen Trends, bis durch ein CUSUM-Kriterium auf Basis der kumulierten Residuen entschieden wird, dass ein neuer Trend angepasst werden muss. Im Unterschied zur robusten Signalextraktion aus Abschnitt 2.2 und der von Mäki-virta et al. (1991) vorgeschlagenen Median-Filterung erhält man mit dieser Methode kein glattes, stetiges Signal, sondern eine Aneinanderreihung linearer Trends, zwischen denen der Übergang auch durch einen Sprung erfolgen kann. Ein auf der Online-Segmentierung aufbauendes Alarmsystem (Charbonnier und Gentil, 2007) basiert auf einem Vergleich der angepassten Trendwerte mit den eingestellten Alarmgrenzen, wobei aber ein Alarm nur ausgelöst wird, wenn das Originalsignal ebenfalls außerhalb der Grenzen liegt. Die qualitative Information aus der Trendanpassung (steigend, fallend, konstant) wird zusätzlich verwendet, um bestimmte klinische Situationen zu erkennen. Eine Anwendung auf Monitoring-

Daten von acht Patienten mit einer Beobachtungsdauer von 36 Stunden liefert eine Reduktion der falschen Alarme um 33%, wobei alle klinisch relevanten Situationen erkannt werden.

Trotz der erheblichen Forschungsleistungen wurde bisher keiner der Ansätze in ein kommerziell erhältliches Monitoring-Gerät integriert. Ein möglicher Grund dafür ist, dass die entwickelten Verfahren meist nur anhand sehr weniger klinischer Daten getestet wurden und somit nicht sicher eingeschätzt werden kann, wie sie sich in der klinischen Praxis verhalten. Daher kann eine sorgfältig geplante klinische Studie, in der neue Alarm-Algorithmen anhand umfangreicher Patientendaten getestet werden, ein erster Schritt zur Umsetzung neuer Verfahren in die klinische Praxis sein.

Kapitel 3

Planung der klinischen Studie

Für eine klinische Studie zur Validierung neuer Alarm-Algorithmen sind verschiedene Szenarien denkbar. Eine Möglichkeit ist eine kontrollierte klinische Studie, bei der eine Patientengruppe mit dem zu validierenden neuen Alarmsystem und eine Kontrollgruppe mit dem standardmäßig verwendeten Alarmsystem überwacht wird. Wenn allerdings vor einer solchen Anwendung nur vage Erkenntnisse darüber vorliegen, wie sicher das neue Verfahren kritische Patientenzustände erkennt, ist eine derartige Vorgehensweise ethisch nicht vertretbar. Eine nicht alarmierte relevante Änderung des Patientenzustandes kann zu lebensbedrohlichen Situationen oder im schlimmsten Fall zum Tod des Patienten führen.

Ein anderer Ansatz ist die parallele Anwendung beider zu vergleichenden Alarmsysteme bei denselben Patienten. Hier ist durch das Standard-Alarmsystem eine gewisse Absicherung gegen die Nicht-Erkennung klinisch relevanter Veränderungen gegeben. Diese Möglichkeit ist jedoch auch nur praktikabel, wenn nicht nur bereits sämtliche Einstellungen des neuen Alarm-Algorithmus spezifiziert sind, sondern dieser auch schon in einem Prototypen implementiert ist. Dieser muss die technische Voraussetzung erfüllen, dass das parallel verwendete Monitoring-Gerät oder das Monitor-Netzwerk der Intensivstation nicht störend beeinflusst wird. Bevor der Aufwand zur Konstruktion eines Prototypen betrieben wird, sollte der potentielle Nutzen des Verfahrens gezeigt worden sein. Dies ist z.B. durch die Erhebung eines umfangreichen und klinisch annotierten Referenzdatensatzes möglich, auf dem neue Alarm-Algorithmen *offline*, d.h. nicht direkt während der Datenaufzeichnung, getestet werden können. Diese Vorgehensweise wird für die in dieser Arbeit beschriebene klinische Studie gewählt, da sich die zu validierenden Verfahren noch in einer frühen klinischen Testphase befinden und kein Prototyp

zur Online-Erprobung existiert. Ort der Datenaufzeichnung ist die Intensivstation für Innere Medizin I des Universitätsklinikums Regensburg. Ausschlaggebend für die Wahl dieses Studienortes ist zum einen die optimale technische Ausstattung mit modernen Monitoring-Geräten und zum anderen die Möglichkeit der Zusammenarbeit mit engagierten Medizinerinnen, die ein großes wissenschaftliches Interesse an der Verbesserung existierender Alarmsysteme zeigen.

Vor Beginn der Studie wurde von den Beteiligten des Online-Monitoring-Projektes an der Technischen Universität Dortmund in Zusammenarbeit mit Medizinerinnen des Klinikums in Regensburg ein Studienprotokoll erarbeitet. In diesem wurden die Studienziele festgelegt, Ein- und Ausschlusskriterien für die Patienten bestimmt, die Datenerhebung geplant und die statistische Auswertungsmethodik festgehalten. Das Protokoll wurde der Ethik-Kommission des Universitätsklinikums Regensburg vorgelegt. Der Antrag wurde im Mai 2004 genehmigt (siehe Anhang A).

Einige der in der Planung vorgesehenen Punkte konnten aus technischen oder organisatorischen Gründen in der Praxis nicht oder nur in eingeschränktem Maße realisiert werden. Darüber hinaus wurden manche Probleme erst während der ersten Datenaufzeichnungen und Probe-Annotationen deutlich, so dass neue Strategien erarbeitet werden mussten. Diese Arbeit ist somit zum Teil auch die Beschreibung des während der Studie erfolgten neuen und andauernden Lösungsprozesses.

Im Folgenden werden die im Studienprotokoll formulierten Ziele der klinischen Studie vorgestellt. Anschließend wird das vorgesehene Patientenkollektiv beschrieben. Es folgt ein Überblick über das Datenerhebungskonzept. Details zur Datenerhebung und Annotation finden sich in Kapitel 4 und 5. Die Grundidee für die statistische Auswertung wird kurz beschrieben, eine ausführliche Darstellung der vorgesehenen statistischen Methoden erfolgt in Kapitel 6.

3.1 Studienziele

Das Hauptziel und der Grund für die Initiierung der klinischen Studie ist die klinische Validierung neuer Alarm-Algorithmen für das intensivmedizinische Online-Monitoring. Es ist der Nachweis zu erbringen, dass ein neuer Alarm-Algorithmus dem einfachen Schwellwertalarmsystem überlegen ist. Genauer soll überprüft werden, ob mit neu entwickelten Verfahren zur Alarmgebung das Auftreten von Fehlalarmen gegenüber den in der Praxis eingesetzten Systemen ohne Sensitivitätsverlust reduziert werden kann. Sensitivität bezieht sich dabei auf das Erkennen „alarmrelevanter Zustände“.

Zusätzlich hat die Studie verschiedene Nebenziele:

Die erhobenen Daten sollen zur Erstellung eines klinisch annotierten Referenzdatensatzes verwendet werden. Dieser liefert die Grundlage für die Offline-Anwendung und Validierung signalextraktionsbasierter Alarm-Algorithmen. Die erfolgreiche Erhebung des Referenzdatensatzes ist demnach die Voraussetzung dafür, dass die für die Erreichung des Hauptziels nötigen Analysen durchgeführt werden können. Der Referenzdatensatz soll zusätzlich auch als Datengrundlage für die Erprobung anderer, in Zukunft noch zu entwickelnder Verfahren, dienen.

Ein weiteres Nebenziel ist die Bestandsaufnahme der Häufigkeit und Relevanz von Alarmen des derzeit eingesetzten Alarmsystems. Die Ergebnisse sollen mit Resultaten aus anderen bisher publizierten Studien zu diesem Thema verglichen werden. Außerdem können auf diese Weise mögliche weitere Schwachstellen des derzeit eingesetzten Alarmsystems aufgedeckt werden, die nicht notwendigerweise nur Schwellwertalarme, sondern auch andere Alarmtypen (z.B. technische Alarme) betreffen.

Ein weiteres im Studienprotokoll formuliertes Nebenziel ist es zu überprüfen, ob mittels neuer statistischer Methoden Alarme zur Früherkennung kritischer Zustände ausgelöst werden können. Dies wird allerdings in dieser Arbeit nur am Rande diskutiert. Das letzte Nebenziel der Studie – die Untersuchung multivariater Verfahren zur Alarmgebung – ist ebenfalls nicht Gegenstand dieser Arbeit. Es umfasst z.B. die Anwendung multivariater Signalextraktionsmethoden für die Alarmgebung oder die Erprobung von Alarmregeln, die mittels Klassifikationsverfahren aus den annotierten Daten generiert bzw. „gelernt“ werden.

3.2 Patientenkollektiv

Die Studienpatienten rekrutieren sich ausschließlich aus den Patienten der Intensivstation für Innere Medizin I am Universitätsklinikum Regensburg. In die Studie eingeschlossen werden Patienten, die älter als 18 Jahre sind, die mit einem Infinity Monitor[®] (Dräger Medical, Lübeck) überwacht werden und bei denen mindestens kontinuierlich Herzfrequenz, invasiver arterieller Blutdruck und Sauerstoffsättigung gemessen werden. Bevorzugt werden Patienten, bei denen möglichst viele hämodynamische Vitalparameter überwacht werden, sowie Patienten, bei denen ein häufiges Vorkommen von Alarmen zu erwarten ist. Von der Studie ausgeschlossen werden Patienten, die das 18. Lebensjahr nicht vollendet haben, Patienten bei denen nur eine Variable gemessen wird, Patienten bei denen der Blutdruck nicht invasiv gemessen wird und Patienten, bei denen alle

Alarmer über einen längeren Zeitraum deaktiviert sind.

Da die Häufigkeit von Alarmen und von alarmrelevanten Zuständen vor Beginn der Studie nicht zuverlässig vorhergesagt werden kann, gibt es keine statistische Fallzahlplanung in Bezug auf den Umfang der Datensammlung. Im Studienprotokoll ist lediglich festgelegt, dass mindestens 100 Patienten eingeschlossen werden sollen. Zum Zeitpunkt der Verfassung des Protokolls war in Bezug auf die klinischen Annotationen noch geplant, dass sie von einer geschulten Person direkt am Patientenbett durchgeführt werden, die jeden Patienten acht Stunden beobachtet. Anvisiert war somit eine Gesamtstundenzahl von mindestens 800 Stunden. Die Annotation durch einen klinischen Beobachter direkt am Patientenbett war jedoch nicht zu realisieren. Stattdessen wurde beschlossen, dass Videoaufzeichnungen des Patienten zu einem späteren Zeitpunkt durch einen Mediziner beurteilt werden. Diese Änderung des ursprünglichen Plans wurde in einem Amendment zum Studienprotokoll festgehalten, welches im Oktober 2004 von der Ethik-Kommission in Regensburg positiv verabschiedet wurde (siehe Anhang A). In diesem Amendment wurde auch bestimmt, dass die Beobachtungsdauer pro Patient von acht auf 24 Stunden verlängert und die Anzahl der Patienten auf 35 reduziert wird. Nach ersten Probeaufzeichnungen zeigte sich jedoch, dass eine lückenlose Datenaufzeichnung eines Patienten über 24 Stunden nur selten möglich ist. Die Patienten auf der Intensivstation werden z.B. häufig für Untersuchungen wie etwa Röntgen in andere Zimmer gebracht, wofür das Monitoring unterbrochen wird. So werden in die Studie auch Patienten eingeschlossen, für die die Datenaufzeichnung Lücken aufweist und für die insgesamt weniger als 24 Stunden Daten vorliegen. Damit insgesamt dennoch die vorher geplante Stundenzahl von 840 (35×24) Stunden erbracht wird, werden mehr als 35 verschiedene Patienten in die Studie eingeschlossen.

3.3 Überblick über das Datenerhebungskonzept

Für die ausgewählten Studienpatienten werden die folgenden Daten erhoben:

- Alle numerischen Messwerte des Patientenmonitors
- Alle Alarmeinstellungen
- Alle Alarme
- Klinische Annotationen
- Demographische Daten

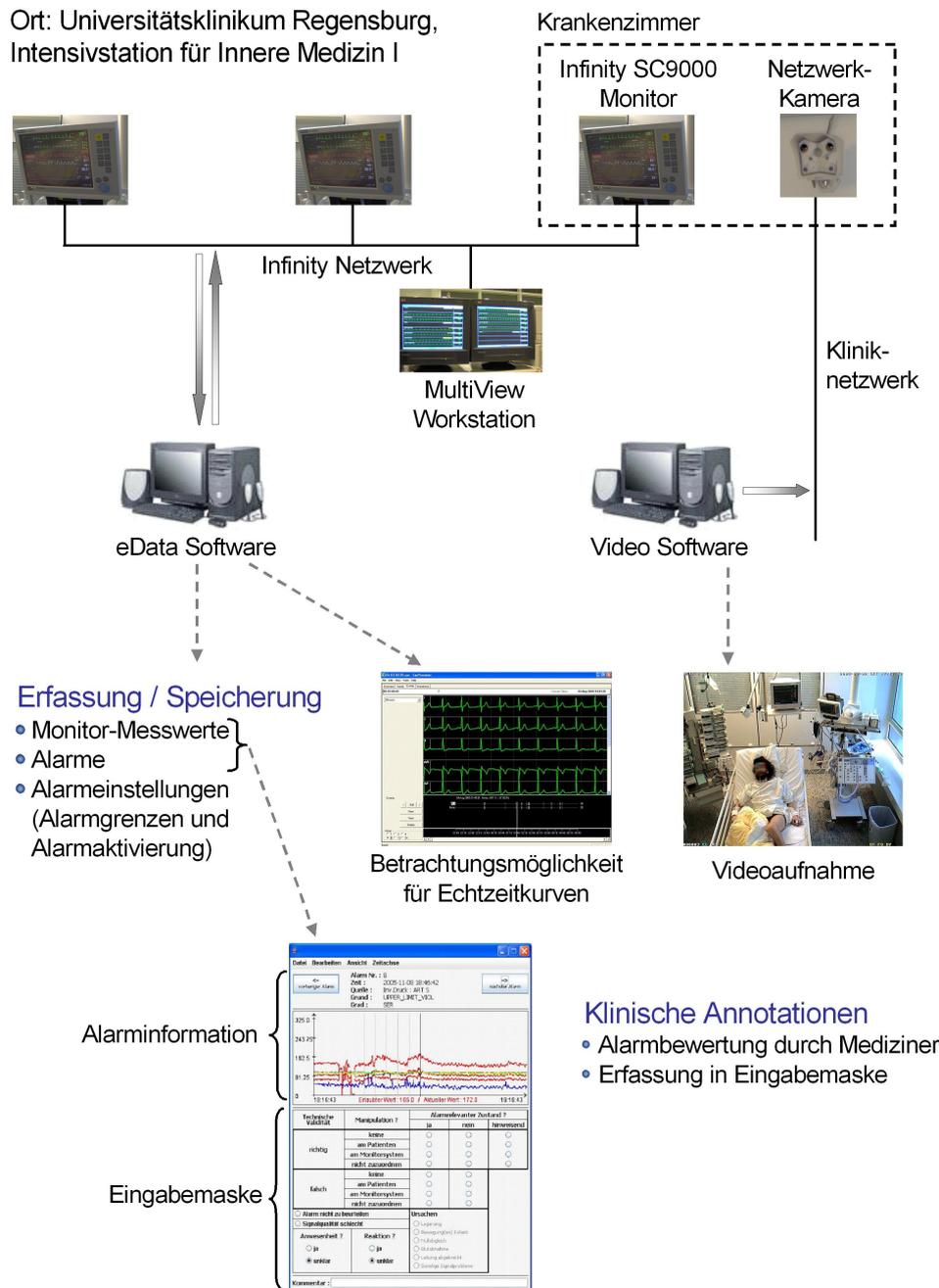


Abbildung 3.1: Schematische Darstellung des Datenerhebungskonzeptes

In Abbildung 3.1 ist das Datenerhebungskonzept schematisch dargestellt. Die numerischen Messwerte, Alarmeinstellungen und Alarme werden sekundengenau und zeitsynchron aus dem Monitoring-Netzwerk extrahiert (siehe Kapitel 4). Klinische Annotationen werden von einem erfahrenen Mediziner auf Basis von Videoaufzeichnungen und Original-Monitordaten vorgenommen und mit einer speziellen Eingabemaske erfasst (siehe Kapitel 5). Die erhobenen Daten aller Studienpatienten bilden den Referenzdatensatz.

Zur Datenerhebung und Annotation ist an dieser Stelle anzumerken, dass sie unter anderem durch technische Probleme immer wieder verzögert wurden. Obwohl nach dem positiven Votum der Ethik-Kommission im Mai 2004 die erste Software-Version für die Aufzeichnung der Daten vom Monitoring-Gerät bereits im August 2004 auf der Intensivstation des Universitätsklinikums in Regensburg installiert wurde, konnte mit der Datenerhebung für den klinisch annotierten Referenzdatensatz erst im Januar 2006 begonnen werden. Bis November 2004 war die Aufzeichnung der Messwerte nur im 10-Sekundentakt und nicht sekundlich möglich. Alarmmeldungen konnten erst im April 2005 verlässlich extrahiert werden. Bis zu diesem Zeitpunkt dauerte auch die Beschaffung und Installation der beiden Kameras für die Videoaufzeichnung. Die Entwicklung der Eingabemaske für die Annotationen war erst im Oktober 2005 vollständig abgeschlossen, da sich nach Probe-Annotationen Änderungsnotwendigkeiten ergaben, wie z.B. die Einbindung einer grafischen Darstellung für die Zeitreihen und Alarme der wichtigsten Vitalparameter. Eine neue Software-Version ermöglichte ab November 2005 die kontinuierliche Aufzeichnung der Alarmgrenzen und im Januar 2006 konnten auch Informationen über die Alarm-Deaktivierungszeiträume erfasst werden, so dass mit der Datenaufzeichnung begonnen werden konnte.

3.4 Planung der statistischen Analyse

Die Vorgehensweise für die statistische Auswertung in Bezug auf das Hauptziel der Studie ist im Studienprotokoll festgelegt. Als Bewertungskriterien für die Güte eines Alarmsystems werden Sensitivität und Spezifität gewählt, welche bekannte Gütemaße aus der Methodik zum Vergleich diagnostischer medizinischer Verfahren sind. Sensitivität und Spezifität bieten eine Möglichkeit die Anforderungen an ein zuverlässiges intensivmedizinisches Alarmsystem zu quantifizieren. Wie sicher alarmrelevante Patientenzustände erkannt werden, kann mittels der Sensitivität beurteilt werden. Eine hohe Spezifität bedeutet, dass in Zeiträumen mit unauffälligem Patientenzustand wenige unerwünschte Alarme ausgelöst werden. Sensitivität und Spezifität werden auch in der Literatur zu Studien über die Bewertung von intensivmedizinischen Alarmsystemen häufig als Gütemaße herangezogen (siehe z.B. Lawless, 1994 und Chambrin et al., 1999).

Für die Validierung eines neuen Alarm-Algorithmus ist in Bezug auf die Sensitivität dessen „Nicht-Unterlegenheit“ gegenüber dem einfachen Schwellwertalarmsystem zu zeigen, wobei die Nicht-Unterlegenheitsschranke von den an der Studie beteiligten Me-

diziniern auf einen allgemein üblichen Wert von 5% festgesetzt wird. Werte von z.B. 7% oder 10% wären in diesem Zusammenhang auch denkbar gewesen. Kann *zusätzlich* die Überlegenheit des neuen Alarm-Algorithmus in Bezug auf die Spezifität gezeigt werden, gilt dies als Nachweis für die Gesamt-Überlegenheit des neuen Verfahrens.

Bei der Planung der Studie wurde bewusst davon abgesehen für den neuen Alarm-Algorithmus die Überlegenheit in Bezug auf die Sensitivität zu fordern. Der ursprüngliche Grund hierfür lag darin, dass für das derzeit eingesetzte Alarmsystem von einer sehr hohen Sensitivität ausgegangen wurde. Die Erfahrung von Medizinern zeigt, dass falsch negative Alarmsystem-Entscheidungen, d.h. vom System nicht erkannte alarmrelevante Zustände, äußerst selten auftreten. Diese Fälle sind dann meist auf zu weit eingestellte Alarmgrenzen zurückzuführen, so dass die Nicht-Alarmierung nur bedingt dem Alarmsystem und vielmehr dem medizinischen Personal anzulasten ist. Auch in den einschlägigen Studien zur Beurteilung von intensivmedizinischen Alarmsystemen wird die Sensitivität selten bemängelt. Da somit für einen neuen Alarm-Algorithmus wenig Aussicht besteht, hinsichtlich der Sensitivität noch eine Verbesserung zu erreichen, wurde hier „nur“ die Nicht-Unterlegenheit gefordert. Die Konzeption der Auswertungsmethodik für das Studienprotokoll erfolgte allerdings unter der Annahme, dass potentielle falsch negative Entscheidungen des derzeit verwendeten Alarmsystems prinzipiell erkannt werden können, da die Annotation durch einen Beobachter direkt am Patientenbett geplant war. Bei der letztendlich umgesetzten Annotation mit Hilfe von Videoaufnahmen ist dies nicht möglich, da das Video im Annotationsprozess immer zur nächsten Alarmsituation vorgespult wird. Dies hat zur Folge, dass für das derzeit eingesetzte Alarmsystem aufgrund des Studiendesigns eine Sensitivität von 100% angenommen wird. Das Nicht-Unterlegenheits-Testproblem zum Vergleich zweier Sensitivitäten kann dann zu einem Testproblem zum Vergleich der Sensitivität des neuen Alarm-Algorithmus mit dem festen Wert 95% vereinfacht werden. Die Annotation nur für einzelne Alarmzeitpunkte zieht außerdem die Notwendigkeit nach sich, dass definiert werden muss, wann ein alarmrelevanter Zustand vom neuen Alarm-Algorithmus erkannt wird, da dieser selten in genau derselben Sekunde alarmiert wie das derzeit eingesetzte System.

Auch die Spezifität ist in Bezug auf Alarmsysteme ein nicht unproblematisches Gütemaß. Hier ist dem Anwender ein großer Definitionsspielraum gegeben, was auch die Ursache dafür ist, dass in den oben erwähnten Studien über Alarmsysteme keine einheitliche Definition für die Spezifität verwendet wird. Im Studienprotokoll wurde eine Vorgehensweise zur Spezifitätsbestimmung auf Basis einer Aufteilung des Beobachtungszeitraums in feste Zeitabschnitte vorgeschlagen. Dieser Ansatz kann in bestimm-

ten Datensituationen zu irreführenden Ergebnissen führen. Da die problematischen Datensituationen in den Daten vom Monitorsystem häufig auftreten, müssen alternative Spezifitätsdefinitionen in Betracht gezogen werden.

Eine ausführliche Darstellung und Diskussion der statistischen Auswertungsmethoden findet sich in Kapitel [6](#).

Kapitel 4

Datenerhebung vom Monitorsystem

Im Folgenden wird zunächst das auf der internistischen Intensivstation des Klinikums der Universität Regensburg eingesetzte Infinity Monitoring System[®] (Dräger Medical, Lübeck) vorgestellt, mit dem die Daten für die Studie erhoben werden. Anschließend erfolgt eine ausführliche Darstellung der Datenerfassung sowie ein Vergleich mit Datenerhebungen anderer Studien.

4.1 Das Infinity Monitoring System[®]

Die hier vorgestellten Informationen zur Funktionsweise des Infinity Monitoring Systems[®] beruhen auf den Angaben in der Bedienungsanleitung (Dräger, 2003).

Das Hauptbild des Monitorbildschirms zeigt standardmäßig Kurven und Parameterfelder. Bei den Kurven handelt es sich z.B. um die Darstellung verschiedener EKG-Ableitungen, um Druckkurven für die verschiedenen invasiv gemessenen Blutdrücke oder um die Sauerstoffsättigungskurve. Die Parameterfelder befinden sich neben den entsprechenden Kurven und enthalten die daraus aktuell errechneten Werte der Vitalparameter (z.B. der Herzfrequenz), die eingestellten Alarmgrenzen und evtl. andere Überwachungssymbole. Es besteht die Möglichkeit, im Hauptbild neben den Kurven auch die Zeitreihen der Vitalparameter für einen Zeitraum von bis zu maximal einer Stunde und für bis zu acht Parameter einzublenden.

4.1.1 Alarme

Alarmgrade

Bei Alarmen werden die drei Alarmgrade „lebensbedrohlich“, „ernst“ und „hinweisend“ unterschieden. Die verschiedenen Alarmgrade werden durch unterschiedliche Alarmtöne und Alarmfarben gekennzeichnet. Auf einer Infozeile am unteren Bildschirmrand wird zudem die Alarmursache eingeblendet. Wenn mehrere Alarme gleichzeitig auftreten, blinken die entsprechenden Parameterfelder, der Alarmton richtet sich aber nach dem Alarm mit dem höchsten Alarmgrad. Für die meisten Alarme kann der Alarmgrad vom Anwender bestimmt werden.

Den höchsten Alarmgrad weisen die lebensbedrohlichen Alarme auf (Abkürzung: LT). Lebensbedrohliche Alarme sind fortbestehend bzw. nicht-selbstlöschend, d.h. die akustischen und visuellen Alarmanzeigen bleiben solange bestehen, bis der Alarm durch das medizinische Personal (mit der „Alarm Stop“- oder der „Alle Alarme Aus“-Taste, s.u.) quittiert wird. Geschieht dies nicht, ertönt der Alarm selbst dann weiter, wenn die Alarmursache in der Zwischenzeit nicht mehr vorliegt. Ein lebensbedrohlicher Alarm tritt z.B. immer bei einer Asystolie oder einer ventrikulären Fibrillation auf.

Ernste Alarme (Abkürzung SER) besitzen den zweithöchsten Alarmgrad. Hier ist nur die *Alarmmeldung* fortbestehend d.h. sie bleibt auch auf der Infozeile eingeblendet wenn die Alarmsituation bereits behoben ist. Dagegen blinkt das Parameterfeld nicht mehr und der Alarmton bricht ab, wenn kein Alarmgrund mehr besteht.

Hinweisende Alarme (Abkürzung ADV) werden bei technischen Problemen (z.B. bei einem abgezogenen Elektrokabel) ausgelöst. Alle Alarmanzeigen hören auf, wenn die Alarmsituation nicht mehr vorliegt oder wenn der Alarm quittiert wird.

Aufhebung der Alarmfunktion

Es besteht die Möglichkeit mit der sogenannten „Alle Alarme Aus“-Taste alle visuellen und akustischen Alarmanzeigen für eine vom klinischen Personal einstellbare Zeit aufzuheben (1, 2, 3, 4 oder 5 Minuten). Eine unbegrenzte Alarmaufhebung ist ebenfalls möglich. Am Studienort Regensburg beträgt die Alarmaufhebungszeit drei Minuten. Die Alarme bleiben solange ausgeschaltet, bis die Taste erneut betätigt wird oder die Frist der Alarmaufhebung abläuft.

Mit der „Alarm Stop“-Taste wird der Alarmton für 60 Sekunden ausgeschaltet, alle visuellen Alarmanzeigen sind aber auf dem Bildschirm zu sehen. Der Alarmton tritt erneut auf, wenn während der 60 Sekunden ein neuer Alarm auftritt oder wenn eine lebensbedrohliche oder ernste Alarmsituation auch nach der Minutenfrist noch vorliegt.

Alarmeinstellungen

Auf einer sogenannten Alarmgrenzwerttabelle kann die Alarmfunktion gezielt für jeden einzelnen Parameter ein- oder ausgeschaltet werden. Weiterhin können die unteren und oberen Alarmgrenzen für die verschiedenen Parameter innerhalb vorgegebener Einstellungsgrenzen individuell eingestellt werden oder die voreingestellte Grund-Alarmeinstellung wird übernommen. Es besteht außerdem die Möglichkeit einer automatischen Grenzwerteinstellung, d.h. die Alarmgrenzen werden relativ zu den Parameterwerten eingestellt, die bei Auswahl dieser Option aktuell beim Patienten vorliegen.

Auslösung von Schwellwertalarmen

Liegt der Messwert eines Parameters, für den die Alarmfunktion eingeschaltet ist, außerhalb des eingestellten Alarmbereichs, ertönt nicht in jedem Fall sofort ein Alarm. Für bestimmte Parameter besteht die Option der sogenannten Alarmbestätigung. Ist diese eingeschaltet, so wird für eine bestimmte Anzahl nachfolgender Parameterwerte geprüft, ob die Alarmsituation immer noch vorliegt, bevor ein Alarm ausgelöst wird. Das Ziel dieser Alarmverzögerung ist die Vermeidung von Fehlalarmen. Tabelle 4.1 enthält die in Regensburg verwendeten Alarmbestätigungsfristen.

Parameter	Obere Alarmgrenze	Untere Alarmgrenze
Herzfrequenz	nach 2 Sek.	sofort
Invasive Blutdrücke	4 Sek.	4 Sek.
Sauerstoffsättigung	4 Sek.	10 Sek.
Atemfrequenz	8 Sek.	10 Sek.
Puls	4 Sek.	10 Sek.
Alle anderen Parameter	sofort	sofort

Tabelle 4.1: Alarmbestätigungsfristen der Parameter für die Überschreitung der oberen Alarmgrenze bzw. für die Unterschreitung der unteren Alarmgrenze

Arrhythmieerkennung

Arrhythmiealarme basieren auf dem EKG und folgen einem anderen Prinzip als Schwellwertalarme. Zunächst erstellt der Monitor in einer 30 bis 40 Sekunden dauernden Lernphase anhand der vorherrschenden Gestalt der EKG-Kurve für einen Herzschlag ein Referenzmuster. Alle darauf folgenden Herzschläge werden mit diesem Muster verglichen und als gültig oder ungültig klassifiziert (Schlagklassifizierung). Bei der Rhythmusklassifizierung werden dann *Folgen* gültiger Schläge mit gespeicherten Schlagfolgen-Mustern verglichen. Anhand festgelegter Regeln wird festgestellt, ob und wenn ja, welche Arrhythmie vorliegt.

4.1.2 Häufig überwachte Vitalparameter

Herzfrequenz

Die Herzfrequenz (HR) wird mit Hilfe des EKGs gemessen. Für die Berechnung der Herzfrequenz werden die Abstände zwischen den Herzschlägen der letzten zehn Sekunden betrachtet. Aus diesen Abständen wird der Durchschnitt bestimmt, wobei die zwei längsten und die zwei kürzesten Abstände von der Berechnung ausgeschlossen werden. Das Ergebnis wird so transformiert, dass es die Anzahl der Herzschläge pro Minute angibt.

Invasiv gemessene Blutdrücke

Invasive Blutdruckmessung bedeutet, dass die Blutdruckmessung „blutig“ erfolgt, indem ein Katheter in den Blutkreislauf des Patienten eingeführt wird. Tabelle 4.2 gibt eine Übersicht darüber, welche Blutdrücke vom Monitoring-System invasiv überwacht werden können.

Parameter	Druck	Gemessene Werte
ART	Arteriell	Systolisch (ART.S), Diastolisch (ART.D), Mittel (ART.M)
PA	Pulmonalarteriell	Systolisch (PA.S), Diastolisch (PA.D), Mittel (PA.M)
CVP	Zentralvenös	Mittel
LV	Linksventrikulär	Systolisch (LV.S), Diastolisch (LV.D), Mittel (LV.M)
RV	Rechtsventrikulär	Systolisch (RV.S), Diastolisch (RV.D), Mittel (RV.M)
RA	Rechtsatrial	Mittel
LA	Linksatrial	Mittel
ICP	Intrakranial	Mittel

Tabelle 4.2: *Blutdruck-Parameter*

Sauerstoffsättigung

Der Parameter Sauerstoffsättigung (SpO₂) gibt den Prozentsatz von sauerstoffreichem Hämoglobin im arteriellen Blut des Patienten an und wird mit Hilfe der Pulsoxymetrie bestimmt. Ein Lichtsensor am Finger, Zeh oder Ohrläppchen des Patienten misst die Absorptionsgrade von rotem und infrarotem Licht. Sauerstoffreiches und sauerstoffarmes Blut absorbieren rotes Licht unterschiedlich, infrarotes Licht jedoch fast gleich. Aus den unterschiedlichen Absorptionsgraden lässt sich daher der Anteil von sauerstoffreichem Blut errechnen. Die Sauerstoffsättigung ändert sich zyklisch mit dem Blutvolumen. Da das Blutvolumen von der Pulsrate abhängt, kann über die SpO₂-Überwachung auch die Pulsrate (PLS) bestimmt werden.

Atemfrequenz

Die Überwachung der Atemfrequenz (RESP) basiert auf der Impedanzmethode. Hierbei wird zwischen den EKG-Elektroden auf der Brust des Patienten ein harmloser hochfrequenter Strom erzeugt. Mit Ausdehnung und Erschlaffung des Brustkorbes während des Einatmens und Ausatmens ändert sich der elektrische Widerstand (Impedanz) zwischen den Elektroden. Die Atemfrequenzkurve und die Atemfrequenz werden anhand dieser zyklischen Widerstandsänderungen bestimmt.

Nicht-invasiv gemessene Blutdrücke

Blutdrücke können auch nicht-invasiv über eine Blutdruckmanschette gemessen werden. Die Messungen für systolischen (NBP.S), diastolischen (NBP.D) und mittleren (NBP.M) nicht-invasiven Blutdruck sind nicht kontinuierlich, sondern erfolgen in größeren Abständen, z.B. alle 15 Minuten.

4.2 Die Datenerfassung

Die Monitordaten werden mit einer speziellen Software (eData[®], Dräger Medical, Lübeck) in Echtzeit aus dem Infinity[®]-Netzwerk (Dräger Medical, Lübeck) abgerufen und in full-disclosure-Dateien gespeichert. Dabei werden alle Daten vom bettseitigen Monitor mit einem Zeitlabel versehen. Das Monitor-Netzwerk, die bettseitigen Monitore und alle für die Datenerhebung genutzten Computer sind mit dem auf deutsche Standardzeit (PTB-Zeit) eingestellten Klinik-Netzwerk synchronisiert.

Messwerte

Alle numerischen Messwerte der überwachten Vitalparameter werden mit Hilfe der eData[®]-Software im Sekundentakt aus den full-disclosure-Dateien extrahiert und im Textformat gespeichert. Diese Daten bilden die Entscheidungsgrundlage des derzeitigen Schwellwertalarmsystems und dienen auch als Input für die Offline-Anwendung neuer Alarm-Algorithmen.

Alarmeinstellungen

Die zu Beginn der Datenaufzeichnung eingestellten Alarmgrenzen sowie jede Veränderung werden mit sekundengenauen Zeitangaben ebenfalls aus den full-disclosure-Dateien extrahiert. Die Kenntnis der Alarmgrenzen ist von großer Wichtigkeit für die Offline-Anwendung des auf der Online-Signalextraktion basierenden Alarm-Algorithmus und dessen Vergleich mit dem einfachen Schwellwertalarmsystem. Da

beide Verfahren Alarmgrenzen verwenden, ist ein Vergleich der Methoden nur sinnvoll, wenn der zu validierende Alarm-Algorithmus mit denselben Grenzen angewendet wird, die auch tatsächlich während der Datenerhebung gültig waren. Für die Wahl der Alarmgrenzen wurde in dieser Studie keine feste Vorgehensweise vorgegeben. Das medizinische Personal kann die Alarmgrenzen nach eigenem Ermessen einstellen und verändern.

Weiterhin ist für jeden Vitalparameter bekannt, wann die Alarmfunktion gezielt für diesen Parameter ein- oder ausgeschaltet wurde. Die Speicherung erfolgt zusammen mit den Alarmgrenzen in einer Textdatei. Alarmaufhebungen für alle Alarme mittels der „Alle Alarme Aus“-Taste werden ebenfalls registriert und in einer separaten Textdatei abgelegt. Diese Informationen sind für die Offline-Anwendung der neuen Prozedur ebenfalls von Bedeutung, da in Alarm-Deaktivierungszeiträumen spezielle Regeln für die Auslösung neu generierter Alarme gelten müssen, um einen fairen Vergleich sicherstellen zu können (siehe Kapitel 8).

Alarme

Auch alle Alarme des Infinity Monitoring Systems[®] und die genauen Zeitpunkte ihres Auftretens werden mittels der eData[®]-Software aus den full-disclosure-Dateien gewonnen und in einer Textdatei gespeichert. Die Alarminformation besteht aus dem Alarmgrad, dem alarmlösenden Parameter und der Alarmmeldung. Im Falle eines Schwellwertalarms ist auch die verletzte Alarmgrenze angegeben.

4.3 Datenerhebungskonzepte anderer Studien

In verschiedenen Publikationen wird über die Erhebung intensivmedizinischer Monitor-daten berichtet. Hier erfolgte die Datenerfassung meist am Patientenbett durch einen direkt an den Patientenmonitor angeschlossenen PC (Tsien und Fackler, 1997, Chambrin et al., 2003), während in der hier vorgestellten Studie mit einem PC außerhalb des Krankenzimmers direkt auf das Infinity[®]-Netzwerk zugegriffen wird. Durch den Netzwerkzugriff können nicht nur die Daten eines einzelnen Monitors, sondern bei Bedarf aller an das Netzwerk angeschlossenen Monitore erfasst werden. Neben Daten von Monitoren aus dem mit Kameras ausgestatteten Zimmer können so zwecks Repräsentativitätsüberprüfung auch große Mengen Monitor-daten aus anderen Zimmern erhoben werden.

Während bei der vorliegenden Studie numerische Messwerte, Alarmeinstellungen und

Alarme automatisch mittels der eData[®]-Software extrahiert werden, wurden in anderen Studien nur die numerischen Messwerte automatisch erfasst. Alarmeinstellungen und Alarme wurden manuell als Teil der Annotationen erhoben. Die Erfassung der numerischen Messwerte erfolgte bei Tsien und Fackler (1997) alle fünf bis sechs Sekunden, bei Chambrin et al. (2003) teilweise sekundlich und teilweise alle fünf Sekunden. In der vorliegenden Studie erfolgt die Erhebung der Messwerte im Sekundentakt und die einzelnen Informationen können einander exakt zugeordnet werden, so dass insgesamt eine verlässlichere Datenaufzeichnung als in anderen Studien möglich ist.

Kapitel 5

Klinische Annotationen

Die klinischen Annotationen schaffen die Voraussetzungen für den Vergleich unterschiedlicher Alarmsysteme, da sie Informationen über den wahren klinischen Hintergrund während der Datenaufzeichnung liefern, die aus den Monitordaten allein nicht ersichtlich sind. In der vorliegenden Studie beziehen sich die Annotationen auf die Alarmzeitpunkte des Infinity Monitoring Systems[®]. Im Folgenden wird darauf eingegangen, welche Hilfsmittel bei der Bewertung zum Einsatz kommen und nach welchem Prinzip die Alarme beurteilt werden. Schließlich werden Annotationskonzepte anderer Studien vorgestellt und mit dem Konzept der vorliegenden Studie verglichen.

5.1 Hilfsmittel

5.1.1 Videoaufzeichnung

Während der Datenaufzeichnung werden Videoaufnahmen erstellt, die bei späterer Annotation betrachtet werden. Die Aufzeichnung erfolgt mittels einer Netzwerkkamera (Mobotix M10D, DualNight[®]), welche an der dem Patientenbett gegenüberliegenden Wand angebracht ist. Am Studienort Regensburg sind zwei Bettplätze mit einer solchen Kamera ausgestattet, wobei sich die Betten in demselben Zimmer befinden. Auf der Videoaufnahme sind der Patient selbst und die ihn umgebenden Überwachungsgeräte sichtbar. Pflegehandlungen können ebenfalls beobachtet werden. Bei Dunkelheit schaltet die Kamera automatisch auf Nachtmodus um. Am Bildrand der Videoaufzeichnung sind Datum und Uhrzeit eingeblendet, so dass das Video beim späteren Betrachten



Abbildung 5.1: Videoaufzeichnung

zum gewünschten Alarmzeitpunkt gespult werden kann. Da der Bildschirm des Monitorgeräts auf dem Video sichtbar ist und Alarme auf dem Monitorbildschirm durch ein Aufleuchten des entsprechenden Parameterfeldes signalisiert werden, kann das Auftreten eines Alarms im Video genau beobachtet werden (siehe [Abbildung 5.1](#)).

Die Videos sind nur den für die Annotationen zuständigen Personen zugänglich und werden nach der Auswertung vernichtet. Die Patienten und ihre Verwandten werden durch ein Informationsblatt über die Studie und die zwischenzeitliche Erfassung persönlicher Daten mittels der Kamera informiert (siehe [Anhang A](#)). Eine Einverständniserklärung der Patienten muss laut Beschluss der Ethik-Kommission nicht eingeholt werden.

5.1.2 Eingabemaske

Zur Erfassung der Annotationen wurde eine Eingabemaske entwickelt. Die Programmierung der Maske erfolgte in JAVA. Im Verlauf der Studie wurde die Maske kontinuierlich verbessert, erweitert und den Bedürfnissen der annotierenden Medizinerin optimal angepasst. In ihrer endgültigen Form bietet sie neben einer übersichtlichen Eingabeoberfläche für die Annotationen auch eine zusätzlich eingefügte grafische Aufbereitung der sekundlichen Messwerte der Vitalparameter, welche eine wichtige zusätzliche Entscheidungshilfe für die Alarmbewertung darstellt. Ein Screenshot der Maske findet sich in [Abbildung 5.2](#). Im oberen Teil der Maske ist die für einen Alarm verfügbare Information dargestellt. In den unteren Teil der Maske kann die Alarmbewertung eingegeben werden. Für die Darstellung der Alarminformation müssen vor Beginn der Annotation zunächst zwei Dateien eingelesen werden, welche die Alarmmeldungen und die

Datei Bearbeiten Ansicht Zeitachse
 Alarm Nr. : 234 (verbleibend 346 / 580)
 Zeit : 2006-02-23 15:05:58
 Quelle : Inv.Druck : ART S
 Grund : UPPER_LIMIT_VIOL
 Grad : SER

Erlaubter Wert : 169.0 / Aktueller Wert : 170.0

Technische Validität	Manipulation ?	Alarmrelevanter Zustand ?		
		ja	nein	hinweisend
richtig	keine	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
	am Patienten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	am Monitorsystem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	nicht zuzuordnen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
falsch	keine	<input type="radio"/>	<input type="radio"/>	
	am Patienten	<input type="radio"/>	<input type="radio"/>	
	am Monitorsystem	<input type="radio"/>	<input type="radio"/>	
	nicht zuzuordnen	<input type="radio"/>	<input type="radio"/>	
<input type="radio"/> Alarm nicht zu beurteilen		Ursachen <input type="radio"/> Lagerung <input type="radio"/> Bewegung(en) Patient <input type="radio"/> Nullabgleich <input type="radio"/> Blutabnahme <input type="radio"/> Leitung abgeknickt <input type="radio"/> Sonstige Signalprobleme		
<input type="radio"/> Signalqualität schlecht				
Anwesenheit ? <input checked="" type="radio"/> ja <input type="radio"/> unklar	Reaktion ? <input checked="" type="radio"/> ja <input type="radio"/> unklar			

Kommentar : Antihypertensive Medikation hoch gestellt

Abbildung 5.2: Eingabemaske

Messwerte der Vitalparameter enthalten.

Die Alarmdatei enthält für jeden Alarm Informationen zum Alarmgrad, zur Alarmmeldung und zum alarmlösenden Parameter zusammen mit der sekundengenauen Angabe der Uhrzeit. Die Alarme werden der Reihenfolge nach nummeriert und die zugehörige Nummer wird im oberen Teil der Maske angezeigt (Alarm Nr.). Zudem wird die Gesamtanzahl der Alarme für den aufgezeichneten Zeitraum angegeben sowie die Anzahl der noch zu annotierenden Alarme. Zur weiteren Information werden ebenfalls das Datum und die genaue Uhrzeit des Alarms aus der Alarmdatei übernommen. Dies ist für die Abstimmung mit der Uhrzeit in der Videoaufnahme nötig. Alarme werden anhand des auslösenden Parameters verschiedenen Quellen zugeordnet. Mögliche Quellen sind: EKG, HR (Herzfrequenz), Inv. Druck (invasiver Blutdruck), SpO2

(Sauerstoffsättigung) und NIBP (nicht invasiv gemessener Blutdruck). Die Quelle eines Alarms findet sich ebenfalls im oberen Teil der Maske. Weiterhin ist die Alarmmeldung angegeben. Diese zeigt z.B. an, ob es sich um einen Schwellwertalarm handelt. Zur Vervollständigung der Information über einen Alarm wird auch der Alarmgrad aufgeführt. Die Datei mit den Messwerten der sekundlich gemessenen Vitalparameter wird zur Erstellung von Grafiken benötigt. Für jeden Alarm aus der Alarmdatei wird standardmäßig eine Grafik erzeugt, in der die Messwerte der sechs Parameter Herzfrequenz (grün), Puls (gelb), systolischer arterieller Blutdruck, arterieller Mitteldruck, diastolischer arterieller Blutdruck (alle rot), und Sauerstoffsättigung (schwarz) in einem Intervall von +/- 30 Minuten um den Alarm als Zeitreihen dargestellt sind. Liegen Messwerte für die Atemfrequenz vor, kann die Zeitreihe dieses Parameters ebenfalls gezeichnet werden (blau). Der Alarmzeitpunkt wird dabei in der Grafik als schwarze senkrechte Linie gekennzeichnet. Treten in dem gezeichneten Intervall weitere Alarme auf, werden diese als hellgraue senkrechte Linien dargestellt. Die Grafiken können jederzeit nach Bedarf modifiziert werden. So kann das Zeichnen bestimmter Parameter unterdrückt und die Größe des gezeichneten Zeitintervalls geändert werden. Im Falle eines Schwellwertalarms für einen der sechs grafisch darstellbaren Parameter erscheinen unter der Zeitachse der Grafik zusätzliche Informationen. Zum einen wird der über- bzw. unterschrittene Schwellwert angezeigt („Erlaubter Wert“), welcher in der Alarmdatei zur Verfügung steht. Zum anderen wird der zum Alarmzeitpunkt vorliegende Messwert des alarmlösenden Parameters angegeben („Aktueller Wert“). Nähere Informationen zur Funktionsweise der Eingabemaske finden sich in Kuhls et al. (2006).

5.1.3 Echtzeitkurven

Wichtige Entscheidungshilfen für die Alarmbewertung stellen nicht nur die sekundlichen Messwerte sondern auch die Echtzeitkurven dar, zu denen die verschiedenen EKG-Ableitungen sowie Druckpuls- und Atemkurven gehören. Das Aussehen der verschiedenen Echtzeitkurven, z.B. die Morphologie einzelner Herzschläge im EKG, liefert viele zusätzliche und klinisch bedeutsame Informationen über die Alarmsituation. Auf dem Monitorbildschirm des Infinity Monitoring Systems[®] sind diese Kurven in der Regel dargestellt und somit auch auf der Videoaufzeichnung zu sehen. Für eine genauere Analyse stehen grafische Darstellungen dieser Kurven in einem speziellen Programm zur Verfügung. Ein Screenshot des Programms findet sich in Abbildung 5.3.



Abbildung 5.3: Echtzeitkurven

5.2 Bewertung von Alarmsituationen

Jeder während der Datenaufzeichnung aufgetretene Alarm des Infinity Monitoring Systems[®] wird von einer erfahrenen Medizinerin auf Basis der in den Videoaufzeichnungen und Monitordaten vorliegenden Informationen bewertet. Die Bewertung erfolgt dabei auf drei unterschiedlichen Ebenen. Zunächst wird die technische Validität des Alarms beurteilt. Ein Alarm wird als *technisch richtig* bewertet, falls

- der Alarm aufgrund eines Messwertes gegeben wird, der dem beim Patienten vorliegenden Wert entspricht oder
- ein technischer Alarm gegeben wird und durch die Alarmmeldung das vorliegende Problem korrekt identifiziert wird.

Dagegen wird ein Alarm als *technisch falsch* beurteilt, falls

- der Alarm aufgrund eines Messwertes gegeben wird, der nicht dem beim Patienten vorliegenden Wert entspricht,
- ein technischer Alarm gegeben wird, ohne dass die Alarmmeldung das vorliegende Problem korrekt identifiziert oder
- ein technischer Alarm gegeben wird, ohne dass überhaupt ein Problem vorliegt.

Weiterhin wird notiert, ob der betrachtete Alarm vom medizinischen Personal durch eine Manipulation herbeigeführt wurde, wobei hier noch zwischen einer Manipulation am Patienten und einer Manipulation am Monitorsystem unterschieden wird.

Um für die spätere Auswertung die Richtigkeit der Alarme des Monitorsystems beurteilen zu können und um Zeiträume zu identifizieren, zu denen die Auslösung eines Alarms nötig ist, muss der medizinische Beobachter entscheiden, ob zum Zeitpunkt des Alarms ein sogenannter *alarmrelevanter Zustand* vorlag oder nicht. Ein solcher ist in der vorliegenden Studie definiert als ein Zustand, auf den

- eine diagnostische oder therapeutische Entscheidung erfolgt oder
- der die Behebung eines dauerhaften technischen Problems erforderlich macht.

Zusätzlich besteht die Möglichkeit einen technisch richtigen Alarm als *hinweisend* zu bewerten, wenn er noch nicht unmittelbar eine Reaktion erfordert, den Arzt aber dazu veranlasst den Patienten ab nun genauer zu beobachten.

Neben der Beurteilung hinsichtlich technischer Validität, Manipulation und Alarmrelevanz werden in den Annotationen weitere Begleitumstände eines Alarms erfasst. Kann die Ursache für einen technisch falschen Alarm identifiziert werden, wird diese mittels der Eingabemaske dokumentiert (siehe Tabelle 5.1). Wenn auf dem Video zu erkennen ist, dass zum Zeitpunkt des Alarms eine Pflegekraft anwesend war oder dass auf den Alarm eine Reaktion erfolgt, wird dies ebenfalls vermerkt. Ist aufgrund von Vorinformationen über den Patienten bekannt, dass die Qualität bestimmter Signale fragwürdig ist, wird dies in den Annotationen festgehalten. Weitere wichtige Beobachtungen und Bemerkungen zu einem Alarm können als Freitext in einem Kommentarfeld der Eingabemaske eingetragen werden. In seltenen Fällen kann es vorkommen, dass ein Alarm mit den vorliegenden Informationen nicht beurteilt werden kann.

Alle Alarme, die sich nach dem vorgestellten Prinzip beurteilen lassen, können gemäß des Schemas in Tabelle 5.2 in *richtige*, *hinweisende*, *falsche* und *herbeigeführte* Alarme klassifiziert werden. Ein technisch richtiger Alarm, der während eines alarmrelevanten Zustands auftritt, ist in jedem Fall ein richtiger Alarm. Bei einem technisch falschen Alarm während eines alarmrelevanten Zustands gestaltet sich die Situation schwieriger: Der Alarm selbst ist nötig, z.B. um auf ein permanentes technisches Problem aufmerk-

Ursache	EKG	HR	Inv. Druck	SpO2	NIBP
Lagerung	X	X	X	X	X
Bewegung(en) Patient	X	X	X	X	X
Sonst. Signalprobleme	X	X	X	X	X
Elektrode ab	X	X			
Niedervoltage	X				
EKG zählt falsch		X			
Nullabgleich			X		
Blutabnahme			X		
Leitung abgeknickt			X		
Dislokation				X	
Perfusionsstörung				X	
Fehlerhafte Platzierung					X

Tabelle 5.1: Mögliche Ursachen für technisch falsche Alarme nach Alarmquellen

Technische Validität	Manipulation	Alarmrelevanter Zustand?		
		ja	nein	hinweisend
richtig	nein	<i>richtig</i>	<i>falsch</i>	<i>hinweisend</i>
	ja		<i>herbeigeführt</i>	
falsch	nein	Alarm nötig, aber	<i>falsch</i>	
	ja	Meldung inadäquat	<i>herbeigeführt</i>	

Tabelle 5.2: Klassifizierung der Alarme des Monitorsystems

sam zu machen, der Alarmtyp und die Alarmmeldung sind allerdings falsch bzw. der Situation nicht angemessen. Alle Alarme, die auftreten, ohne dass ein alarmrelevanter Zustand vorliegt und die nicht als hinweisend bewertet wurden, sind falsche Alarme. Allerdings sind falsche Alarme, die durch eine Manipulation seitens des Pflegepersonals ausgelöst wurden, gesondert zu behandeln. Da diese Alarme aktiv provoziert wurden und somit nicht primär dem Alarmsystem anzulasten sind, werden sie als herbeigeführt bezeichnet.

Eine solch differenzierte Bewertung von offline generierten Alarmen neuer Alarm-Algorithmien ist auf Basis der Annotationen nicht möglich. Die für die Validierung benötigten Gütemaße Sensitivität bezüglich der Erkennung alarmrelevanter Zustände und Spezifität lassen sich aber auch für neue Alarmsysteme bestimmen. Für die Sensitivitätsberechnung sind dabei, ausgehend von den *Alarmzeitpunkten* des derzeit verwendeten Alarmsystems, für die ein alarmrelevanter Zustand festgestellt wurde, alarmrelevante *Zeiträume* zu definieren. Bei der Spezifitätsschätzung besteht die Herausforderung in der Bestimmung der Anzahl nicht alarmrelevanter Zustände und der Anzahl richtig negativer Alarmsystem-Entscheidungen. Eine Beschreibung der genauen Vorgehensweise findet sich in Abschnitt 6.2.

Die Tatsache, dass Annotationen nur für die Alarmzeitpunkte des Infinity Monitoring Systems[®] vorgenommen werden, führt dazu, dass alarmrelevante Situationen, in denen kein Alarm auftritt, nicht identifiziert werden. Aber selbst bei einer vollständigen Betrachtung der Videoaufzeichnungen wäre die Erkennung eines alarmrelevanten Zustandes ohne begleitenden Alarm sehr schwierig. Die klinische Erfahrung zeigt, dass solche falsch negativen Entscheidungen des Alarmsystems selten auftreten und meist ihre Ursache darin haben, dass die Alarmgrenzen zu weit eingestellt sind. In bisherigen Studien zur Häufigkeit und Relevanz von Alarmen wird dies bestätigt. Tsien und Fackler (1997) identifizieren keine einzige falsch negative Situation in 298 Beobachtungsstunden. Chambrin et al. (1999) beobachten 24 solcher Situationen in einem Zeitraum von 1971 Stunden – allerdings wird in dieser Studie sogar eine Situation,

in der eine Verstellung der Alarmgrenzen ohne vorherigen Alarm erfolgt, als falsch negative Entscheidung des Alarmsystems gewertet.

5.3 Annotationskonzepte anderer Studien

In mehreren Beobachtungsstudien wurde bereits die Häufigkeit von Alarmen auf Intensivstationen untersucht und das jeweils verwendete Alarmsystem bewertet.

Ein Unterschied zur vorliegenden Studie besteht darin, dass die Annotationen bei allen Studien nicht auf Basis von Videoaufnahmen vorgenommen werden, sondern durch einen Beobachter direkt am Patientenbett. Bei Koski et al. (1990) und Lawless (1994) erfolgt die Klassifizierung der Alarme durch das Pflegepersonal während der normalen Arbeitszeit. In Studien von Tsien und Fackler (1997), Chambrin et al. (1999), Biot et al. (2000) und Zhang (2003) wird die Alarmbewertung von speziell dafür ausgebildeten Beobachtern durchgeführt, die nicht für die Pflege des Patienten zuständig sind. Ein Vorteil der Verwendung von Videoaufnahmen ist, dass die zu beurteilenden Situationen mehrmals betrachtet werden können, was besonders hilfreich ist, wenn viele Alarme schnell hintereinander auftreten. So können Video und Monitordaten sorgfältig verglichen werden, ohne andere Situationen zu verpassen. Durch einen Beobachter am Patientenbett kann zudem möglicherweise ein verzerrtes Verhalten des Pflegepersonals oder des Patienten bewirkt werden. Weitere Vorteile von Videoaufnahmen sind, dass die Kameras auch nachts im abgedunkelten Zimmer gute Bilder liefern und dass Aufzeichnungen über längere Zeiträume (z.B. 24 Stunden) möglich sind. Möglicherweise nachteilig bei der Annotation mit Hilfe von Videoaufnahmen ist, dass keine Rücksprache mit dem Pflegepersonal bezüglich der Bewertung einzelner Situationen gehalten werden kann. Aber auch eine solche Rücksprache könnte eine Verzerrung bewirken.

In den genannten Studien wird in den Annotationen eine direkte Bewertung der aufgetretenen Alarme vorgenommen. Bei der hier vorgestellten Studie bezieht sich die Bewertung nur teilweise auf den Alarm selbst, etwa bei der Beurteilung der technischen Validität. Die Bewertung hinsichtlich der Alarmrelevanz bezieht sich dagegen auf die *Situation* während des Alarms, da mittels der Annotationen auch andere Alarm-Algorithmus bewertet werden sollen.

Um die Annotationskonzepte der anderen Studien einander gegenüberzustellen, wird eine Kategorisierung der dort vorgenommenen Bewertungen in richtig und falsch positive Alarme und richtig und falsch negative Alarmsystem-Entscheidungen vorgenommen (siehe Tabelle 5.3). Die verschiedenen Studien weisen vor allem die Gemeinsamkeit

auf, dass ein Alarm als richtig positiv bewertet wird, wenn auf ihn eine Aktion erfolgt. Dahinter steckt die Idee, dass ein Alarm, auf den reagiert wird, berechtigt sein muss. Auch in der vorliegenden Studie ist die Entscheidung über das Vorliegen eines alarmrelevanten Zustandes aktions-abhängig. Allerdings unterscheiden sich die Studien in Bezug darauf, *welche* Aktionen die Alarmbewertung „richtig positiv“ nach sich ziehen. Koski et al. (1990) untersuchen in einer Studie die Relevanz und Häufigkeit von Alarmen auf einer Intensivstation für erwachsene Patienten, die am Herzen operiert wurden. In dieser Studie wird ein richtig positiver Alarm als *signifikant* bezeichnet, wobei hier unterschieden wird zwischen einem Alarm, bei dem der Patientenzustand nur untersucht wird (*significant alarm - patient condition checked*) und einem Alarm, der eine therapeutische Handlung zur Folge hat (*significant alarm - treatment implemented*). Alarme aufgrund von technischen Problemen, unabhängig davon, ob das Problem behoben werden muss, um die Patientenüberwachung weiterhin zu gewährleisten, werden als sogenannte *Artefakt-Alarme* angesehen. Ein Artefakt-Alarm liegt ebenfalls vor, wenn die Alarmgrenzen nur bis zu 15 Sekunden lang überschritten werden. Ein Alarm wird als unangemessen (*undue*) betrachtet, wenn die Alarmgrenze zwar für längere Zeit als 15 Sekunden überschritten wird, aber außer einer kurzen Kontrolle des Monitorsystems und dem Abschalten des Alarms keine anderen Handlungen durchgeführt werden. Für die Auswertung werden die Alarme der letzten beiden Kategorien unter der Bezeichnung „nicht signifikant“ (*non-significant*) zusammengefasst. Die Zeiträume, in denen kein Alarm ausgelöst wird, werden bei Koski et al. nicht betrachtet, so dass auch keine Aussagen über die Identifizierung von richtig oder falsch negativen Entscheidungen des Alarmsystems gemacht werden können.

Lawless (1994) analysiert im Rahmen einer klinischen Studie die Alarme eines Alarmsystems auf einer Kinderintensivstation. Er klassifiziert einen Alarm nur dann als richtig positiv, wenn er eine Therapieänderung nach sich zieht. Andere Alarme, auch solche, die aufgrund von technischen Problemen auftreten, werden als falsch positiv bewertet. Alarme, die auf Manipulationen durch das Personal zurückzuführen sind, werden gesondert als herbeigeführt (*induced*) eingestuft. Die Alarme der letzten beiden Kategorien werden für die Auswertung unter dem Begriff „nicht signifikant“ (*non-significant*) zusammengefasst. Wenn das Alarmsystem keinen Alarm auslöst, obwohl sich der Zustand des Patienten klinisch bedeutsam verändert, wird dies in dieser Studie als falsch negative Beobachtung gewertet. Generell problematisch ist die Bestimmung der Anzahl der richtig negativen Situationen. In dieser Studie wird das Problem gelöst, indem ein potentieller Alarm pro fünf Minuten Überwachungszeit angenommen wird. Für jede fünf Minuten Beobachtungsdauer, in denen der Patientenzustand stabil ist und kein

Alarm auftritt, wird somit eine richtig negative Beobachtung gezählt.

Tsien und Fackler (1997) untersuchen die Häufigkeit und Relevanz von Alarmen auf einer multidisziplinären Kinderintensivstation. Hier werden richtig positive Alarme in klinisch relevante und klinisch irrelevante Alarme unterschieden. Ein Alarm wird als klinisch relevant bewertet, wenn zum einen der Wert des alarmlösenden Parameters korrekt gemessen und die festgelegte Alarmgrenze über(unter-)schritten wird und zum anderen der Zustand des Patienten eine sofortige Handlung erforderlich macht. Als klinisch irrelevante Alarme werden solche identifiziert, bei denen der Parameterwert zwar korrekt gemessen und auch die festgelegte Alarmgrenze über(unter-)schritten wird, sich der Zustand des Patienten aber nicht soweit verschlechtert hat, dass eine pflegerische Handlung notwendig ist. Falsch positiv sind in dieser Studie Alarme, die aufgrund eines falsch gemessenen Datenwertes ausgelöst werden. Wenn der Patientenzustand ein Eingreifen nötig macht, aber kein Alarm ausgelöst wird, wird dies als falsch negative Entscheidung des Alarmsystems betrachtet. Jede Periode ohne Alarm und ohne Alarmgrund von mindestens 10 bis 15 Minuten Länge wird als eine richtig negative Beobachtung gezählt. Die hier gewählte Bewertung ist ein Kompromiss zwischen technischer und klinischer Beurteilung eines Alarmsystems. Alarme aufgrund von korrekt gemessenen und außerhalb der Alarmgrenzen liegenden Werten, werden immer als richtig positiv betrachtet (technisch richtig). Alarme aufgrund von fehlerhaft gemessenen Werten außerhalb der Alarmgrenzen dagegen als falsch positiv. Die klinische Komponente wird insofern einbezogen, dass die richtig positiven Alarme noch in klinisch relevante und klinisch irrelevante Alarme aufgeteilt werden.

Bei einer multizentrischen Beobachtungsstudie von Chambrin et al. (1999) auf fünf Intensivstationen für Erwachsene wird ein Alarm immer dann als richtig positiv beurteilt, wenn als Reaktion auf den Alarm eine therapeutische Handlung vorgenommen wird (z.B. neue Medikamentendosierung, Absaugen), Untersuchungen durchgeführt (z.B. Röntgen, Blutgasanalyse), die Alarmgrenzen modifiziert, oder technische Probleme behoben werden (z.B. Neuanschluss oder Reinigung des Katheters). Erfolgt auf einen Alarm keine derartige Aktion, wird er als falsch positiv eingestuft. Situationen, in denen eine der oben genannten Aktionen ausgeführt wird, ohne dass vorher ein Alarm gegeben wird, erhalten die Bewertung falsch negativ. Für die Bestimmung der Anzahl der richtig negativen Situationen wird in dieser Studie zunächst die mittlere Intervalllänge zwischen zwei Alarmen in Bezug auf die gesamte Studiendauer bestimmt. Pro Intervall dieser Länge wird ein potentieller Alarm gezählt, woraus die Anzahl der richtig negativen Fälle resultiert. Für jeden in der Studie beobachteten Alarm wird zudem festgehalten, ob er aufgrund einer Manipulation durch das Personal, wegen eines

technischen Problems oder vom Patienten selbst ausgelöst wird.

In einer Studie von Biot et al. (2000) zur Alarmrelevanz auf einer vorwiegend chirurgischen Intensivstation für Erwachsene werden für jede Minute des Beobachtungszeitraums Bewertungen nach dem Schema richtig positiv, falsch positiv, richtig negativ und falsch negativ vorgenommen. Tritt in einer Minute ein Alarm auf, bei dem die Messung korrekt und außerhalb der Alarmgrenzen ist, wird dieser als richtig positiv gewertet. Richtig positive Alarme, die eine direkte Konsequenz haben, wie z.B. das Rufen eines Arztes, eine therapeutische Handlung, eine sofortige Diagnose, Aussetzen der Pflege oder eine Veränderung der Behandlung, werden zusätzlich als „richtig positiv mit Konsequenz“ bezeichnet. Treten in einer Minute mehrere Alarme aus unterschiedlichen Quellen auf, so wird jeder dieser Alarme einzeln bewertet. Hält ein Alarm mehrere Minuten an, so wird die Alarmbewertung für jede dieser Minuten beibehalten und notiert. Dies geschieht zur besseren Quantifizierung der Lärmbelästigung. Ein Alarm wird in dieser Studie als falsch positiv bewertet, wenn er aufgrund inkorrektur Messwerte, z.B. aufgrund eines verstopften Katheters oder Problemen mit Messfühlern, ausgelöst wird. Wie in der Studie von Tsien und Fackler (1997) wird hier bei der Bewertung auch ein Kompromiss zwischen technischer und klinischer Beurteilung gewählt. Eine falsch negative Beobachtung wird notiert, wenn kein Alarm erfolgt, obwohl der gemessene Wert außerhalb der Alarmgrenzen liegt (technisches Problem) oder wenn eine Folge von Messwerten als bedeutsam eingestuft, aber aufgrund zu weit eingestellter Alarmgrenzen kein Alarm gegeben wird. Für jede Minute, in denen die Messwerte innerhalb der Alarmgrenzen liegen und kein Alarmgrund vorliegt, wird eine richtig negative Beobachtung festgehalten.

Zhang (2003) untersucht in ihrer Master Thesis die Alarme auf einer Kinderintensivstation. Bei der Alarmbewertung wird in dieser Studie mit einer Ausnahme genauso vorgegangen, wie bei Tsien und Fackler (1997) (s.o.). Zhang betrachtet ein Alarmsystem, welches technische Alarme (hervorgerufen durch z.B. Signalqualitätsprobleme, Geräteausfall, Kalibrierung) durch besondere Alarmtöne kennzeichnet. Solche Alarme sind vergleichbar mit den ADV-Alarmen des Infinity Monitoring Systems[®]. Mit der Begründung, dass auf diese Alarme vom Personal anders reagiert wird, weil die Ursachen für den Alarm bekannt sind, fließen sie in dieser Studie nicht mit in die Bewertung ein.

Da die vorgestellten Studien unterschiedliche Bewertungsschemata verwenden, ist ein Vergleich der Studienergebnisse problematisch. Manche in einer Studie als falsch bewertete Alarme werden in einer anderen Studie als richtig beurteilt. Zudem unterscheiden sich die Methoden zur Bestimmung der Anzahl richtig negativer Beobachtungen.

Studie	Richtig positiv	Falsch positiv	Falsch negativ	Richtig negativ
Koski et al. (1990)	Falls als Reaktion auf den Alarm: - Therapeutische Handlung (<i>significant alarm - treatment implemented</i>) - Patient untersucht (<i>significant alarm - patient condition checked</i>)	<i>Non-significant alarms:</i> - <i>Artifact alarm</i> (Technisches Problem oder Alarmgrenzen nur bis zu 15 Sekunden lang überschritten) - <i>Undue alarm - no action taken</i> (Alarmdauer länger als 15 Sek., aber nur kurze Kontrolle des Monitoringsystems und Abschalten des Alarms)	nicht untersucht	nicht untersucht
Lawless (1994)	Falls als Reaktion auf den Alarm eintretende Therapieänderung erfolgt (<i>significant alarm</i>)	<i>Non-significant alarms:</i> - <i>false alarm:</i> Technisches Problem oder Alarm ohne Intervention - <i>induced alarm:</i> vom Personal herbeigeführter Alarm	Falls kein Alarm auftritt und sich der Zustand des Patienten bedeutsam verändert	Ein potentieller Alarm pro 5 Minuten Überwachungszeit
Chambrin et al. (1999)	Falls als Reaktion auf den Alarm therapeutische Handlung, Untersuchungen, Änderung der Alarmgrenzen oder Behebung technischer Probleme	Falls keine Aktion wie bei „Richtig positiv“ erfolgt	Falls Aktion wie bei „Richtig positiv“ erfolgt, aber kein Alarm gegeben wird	Aus den Daten: Bestimmung der mittleren Länge eines Intervalls zwischen zwei Alarmen. Pro Intervall dieser Länge wird ein richtig negativer Alarm gezählt.
Biot et al. (2000)	Falls Messung korrekt und Wert außerhalb der Alarmgrenzen; - Sonderstellung: Alarme mit direkter Konsequenz (Arzt wird gerufen, therapeutische Handlung, sofortige Diagnose, Aussetzen der Pflege, Therapieänderung)	Falls Alarm aufgrund inkorrekt gemessener (z.B. aufgrund von Problemen mit Messfühlem und Schläuchen, verstopfter Katheter usw.)	Kein Alarm, aber - Messwert außerhalb der Alarmgrenzen (technisches Problem) oder - Folge von Messwerten vom Beobachter als bedeutsam eingestuft (Grenzen zu weit eingestellt)	Jede Minute, in der kein Alarm gegeben wird und kein Grund für einen Alarm vorliegt, erhält die Bewertung richtig negativ.
Tsien und Fackler (1997), Zhang (2003)	Falls Messung korrekt und außerhalb der Alarmgrenzen; Unterscheidung: - <i>True positive clinically relevant</i> (Zustand des Patienten macht sofortige Handlung erforderlich) - <i>True positive clinically irrelevant</i> (Keine pflegerische Handlung notwendig)	Falls Alarm aufgrund eines falsch gemessenen Wertes Bei Zhang (2003): technische Alarme (\rightarrow ADV-Alarme) zwar falsch, bei Auswertung aber nicht berücksichtigt	Falls kein Alarm auftritt, aber Patientenzustand ein Eingreifen nötig macht.	Jede Periode ohne Alarm und ohne Handlungsbedarf von mindestens 10 bis 15 Minuten Länge wird als eine richtig negative Beobachtung gezählt.

Tabelle 5.3: Alarmbewertung in sechs verschiedenen klinischen Beobachtungsstudien

Kapitel 6

Auswertungsmethoden

In diesem Kapitel werden die für die Validierung eines neuen Alarm-Algorithmus vorgesehenen statistischen Methoden vorgestellt. Ziel der klinischen Studie ist der Nachweis, dass das neue Verfahren zur Alarmgebung dem derzeitig eingesetzten einfachen Schwellwertalarmsystem überlegen ist. Hierfür müssen zunächst Bewertungskriterien festgelegt werden:

Ein gutes Alarmsystem erkennt sicher alarmrelevante Situationen und erzeugt gleichzeitig wenige Fehllarme. Das Abschneiden eines Alarmsystems in Bezug auf das Erkennen alarmrelevanter Situationen kann mittels der Sensitivität beurteilt werden, welche häufig zur Bewertung von Diagnoseverfahren in der Medizin verwendet wird. Sie gibt in diesem Kontext die bedingte Wahrscheinlichkeit dafür an, dass eine alarmrelevante Situation durch einen Alarm angezeigt wird

$$\text{Sensitivität} = P(\text{Alarm}|\text{Zustand alarmrelevant}) \quad (6.1)$$

und sollte idealerweise einen Wert nahe 1 annehmen. Die Sensitivität kann im konkreten Fall geschätzt werden durch den Quotienten aus der Anzahl der erkannten alarmrelevanten Situationen und der Anzahl aller aufgetretenen alarmrelevanten Situationen. Zur Beurteilung eines Alarmsystems hinsichtlich der Auslösung von Fehllarmen bietet sich die Spezifität – ein weiteres Gütemaß für Diagnoseverfahren – an. Sie ist gegeben durch

$$\text{Spezifität} = 1 - P(\text{Alarm}|\text{Zustand nicht alarmrelevant}) \quad (6.2)$$

$$= P(\text{kein Alarm}|\text{Zustand nicht alarmrelevant}). \quad (6.3)$$

Da eine niedrige Fehllalarmrate erwünscht ist, sollte auch die Spezifität möglichst einen Wert nahe 1 annehmen. Die Schätzung der Spezifität im Alarmsystem-Kontext erfor-

dert eine Bestimmung der Anzahl der nicht alarmrelevanten Zustände. Dies kann auf unterschiedliche Arten geschehen und wurde auch in der Literatur bisher nicht einheitlich gelöst. Hierzu sei auf Tabelle 5.3 und die unterschiedlichen Vorgehensweisen hinsichtlich der Bestimmung der Anzahl richtig negativer Alarmsystem-Entscheidungen verwiesen. Eine ausführliche Diskussion der Problematik findet sich in Abschnitt 6.2. Nach Auswahl von Sensitivität und Spezifität als Beurteilungskriterien ist es dann naheliegend, für den Überlegenheitsnachweis zu fordern, dass der neue Alarm-Algorithmus sowohl hinsichtlich der Sensitivität als auch in Bezug auf die Spezifität besser als das einfache Schwellwertsystem ist. In dieser Studie kann allerdings ein Überlegenheitsnachweis für die Sensitivität nicht geführt werden, weil die Sensitivität für das einfache Schwellwertalarmsystem – bedingt durch das Studiendesign – gleich 100% ist. Da nur Situationen bewertet werden, in denen vom Infinity Monitor[®] Alarme ausgelöst wurden, können keine falsch negativen Entscheidungen des einfachen Schwellwertsystems identifiziert werden (siehe auch Abschnitt 5.2). Stattdessen soll die „Nicht-Unterlegenheit“ des neuen Systems in Bezug auf die Sensitivität gezeigt werden, d.h. es wird gefordert, dass das neue Alarmsystem *nicht wesentlich schlechter* als das Schwellwertalarmsystem abschneidet. Dies setzt eine Festlegung des medizinisch noch akzeptablen Sensitivitätsverlustes voraus. In dieser Studie wird die noch tolerierbare Differenz der Sensitivitäten auf einen in diesem Zusammenhang durchaus üblichen Wert von 5% festgelegt. Um insgesamt die Überlegenheit des neuen Systems zeigen zu können, wird gleichzeitig zur Nicht-Unterlegenheit in Bezug auf die Sensitivität gefordert, dass das neue Verfahren in Bezug auf die Spezifität überlegen ist. Die statistische Methodik zur Behandlung eines solchen Testproblems wird im folgenden Abschnitt 6.1 vorgestellt. In Abschnitt 6.2 wird dargestellt, wie in dieser Studie Sensitivitäten und Spezifitäten für Alarmsysteme bestimmt werden. Abschnitt 6.3 fasst weitere Beurteilungskriterien für Alarmsysteme zusammen.

6.1 Methoden für die Bewertung und den Vergleich von Diagnoseverfahren

In der Medizin werden häufig diagnostische Verfahren verwendet, um bei Patienten z.B. das Vorliegen einer bestimmten Krankheit oder allgemeiner eines bestimmten Zustandes festzustellen. Als Beispiele seien „Tests“ auf HIV-Antikörper im Blutserum oder Schwangerschaftstests genannt. Die statistische Methodik für die Evaluierung

und Gegenüberstellung von Diagnoseverfahren ist in diesem Kontext von Interesse, da ein Alarmsystem als ein Diagnosewerkzeug aufgefasst werden kann, welches kontinuierlich operiert und entscheidet, ob und wann ein Zustand alarmierungswürdig ist. Weil einem Alarmsystem nur zwei Entscheidungsmöglichkeiten zur Verfügung stehen („Alarm“ oder „kein Alarm“), wird im Folgenden nur auf die Methodik zur Bewertung *binärer* diagnostischer Verfahren eingegangen. Analysemethoden für Verfahren mit mehr als zwei möglichen Entscheidungen (sog. quantitative Diagnostiktests) und viele der in den folgenden Abschnitten beschriebenen Methoden, finden sich z.B. in den Lehrbüchern von Abel (1993), Zhou, Obuchowski und McClish (2002) und Pepe (2003).

Zunächst wird die im Folgenden verwendete Notation eingeführt: Mit D (für Decision) sei das zu beurteilende binäre diagnostische Entscheidungsverfahren bezeichnet. Ist $D = 1$, so liefert das Diagnoseverfahren ein positives Resultat, d.h. es entscheidet, dass der zu diagnostizierende Zustand vorliegt (z.B. „HIV Infektion liegt vor“). $D = 0$ steht entsprechend für ein negatives Ergebnis.

Um die Güte eines Diagnoseverfahrens beurteilen zu können, ist die Kenntnis über den wahren Zustand erforderlich. In der Praxis wird der wahre Zustand häufig durch einen sogenannten Goldstandard bestimmt. Dies kann z.B. ebenfalls ein diagnostisches Verfahren sein, das praktisch fehlerfrei, aber möglicherweise sehr viel aufwändiger als das zu beurteilende Diagnoseverfahren durchzuführen ist. Im Falle der vorliegenden klinischen Studie ist die anstelle eines Goldstandards verwendete Referenz für den „wahren“ vorliegenden Zustand die in den Annotationen festgehaltene Bewertung der Alarmsituationen durch eine erfahrene Medizinerin (siehe Kapitel 5.2). Im Folgenden bezeichne Z den Indikator für den wahren Zustand, welcher durch einen Goldstandard verifiziert werden kann. $Z = 1$ bedeutet, dass bei der Untersuchungseinheit (z.B. dem Patienten) der zu diagnostizierende Zustand vorliegt. $Z = 0$ bedeutet dementsprechend, dass dies nicht der Fall ist.

Ist die Bewertung nur eines einzigen diagnostischen Verfahrens D von Interesse und liegt für alle n Untersuchungseinheiten sowohl der wahre Zustand als auch das Diagnose-Ergebnis vor, werden die Beobachtungen häufig in Form einer Vierfeldertafel wie in Tabelle 6.1 dargestellt. n_{zj} , $z, j \in \{0, 1\}$, geben dabei die Anzahl der Untersuchungseinheiten mit dem wahren Zustand z und der Diagnose j an. Die in Klammern angegebenen Größen P_{zj} , $z, j \in \{0, 1\}$, bezeichnen die zugehörigen Wahrscheinlichkeiten. Durch „+“ wird angezeigt, dass über die entsprechenden Indizes summiert wird. Gemäß dieser Notation ist die Sensitivität des Diagnoseverfahrens D gegeben durch

$$\text{Se}(D) = P(D = 1 | Z = 1) = \frac{P_{11}}{P_{1+}}.$$

	$Z = 1$	$Z = 0$	Summe
$D = 1$	$n_{11} (P_{11})$	$n_{01} (P_{01})$	$n_{+1} (P_{+1})$
$D = 0$	$n_{10} (P_{10})$	$n_{00} (P_{00})$	$n_{+0} (P_{+0})$
Summe	$n_{1+} (P_{1+})$	$n_{0+} (P_{0+})$	$n (1.00)$

Tabelle 6.1: Kontingenztabelle für den Vergleich eines diagnostischen Verfahrens D mit dem wahren Zustand Z . Mit n sind die absoluten Häufigkeiten und mit P die zugehörigen Wahrscheinlichkeiten bezeichnet.

P_{1+} wird dabei auch als Prävalenz des zu diagnostizierenden Zustandes bezeichnet. Die Sensitivität wird geschätzt durch

$$\hat{\text{Se}}(D) = \frac{n_{11}}{n_{1+}}.$$

Entsprechend ist die Spezifität von D definiert durch

$$\text{Sp}(D) = P(D = 0 | Z = 0) = \frac{P_{00}}{P_{0+}}.$$

Sie kann geschätzt werden durch

$$\hat{\text{Sp}}(D) = \frac{n_{00}}{n_{0+}}.$$

Sollen zwei Diagnoseverfahren, z.B. ein in der Praxis verwendetes Standardverfahren D_1 und ein neu zu validierendes Verfahren D_2 verglichen werden, können die Beobachtungen in zwei Tabellen der Gestalt von Tabelle 6.1 dargestellt werden. Häufig werden beide Diagnoseverfahren aber auf dieselbe Gruppe von Patienten angewendet, so dass es sich um verbundene Stichproben handelt. Dieser Fall liegt auch bei den zu vergleichenden Alarm-Algorithmen vor, da sie auf dieselben Patientendaten angewendet werden. In diesem Fall ist eine Darstellung mit Hilfe zweier Symmetrietafeln – eine (2×2) -Tafel für den Fall $Z = 1$ und eine weitere für den Fall $Z = 0$ – aussagekräftiger (siehe Tabelle 6.2). Die Information aus den einfachen Kontingenztabellen (Tabelle 6.1) bleibt in Tabelle 6.2 erhalten, diese Darstellung zeigt aber zusätzlich auf, wie oft die beiden Diagnoseverfahren mit ihren Bewertungen übereinstimmen bzw. wie oft dies nicht der Fall ist. Hier repräsentieren n_{zjk} , $z, j, k \in \{0, 1\}$ die Anzahl der Untersuchungseinheiten mit dem wahren Zustand z und den Diagnose-Ergebnissen (j, k) . Die Größen P_{zjk} , $z, j, k \in \{0, 1\}$ bezeichnen wieder entsprechend die zugehörigen Wahrscheinlichkeiten. Mit $p = P_{1++}$ sei die Prävalenz des zu diagnostizierenden Zustandes notiert. Es sei zudem $P_{jk|z} = P(D_1 = j, D_2 = k | Z = z)$ die bedingte Wahrscheinlichkeit der Diagnosen (j, k) bei gegebenem wahren Zustand z .

Z = 1			
	$D_2 = 1$	$D_2 = 0$	Summe
$D_1 = 1$	$n_{111} (P_{111})$	$n_{110} (P_{110})$	$n_{11+} (P_{11+})$
$D_1 = 0$	$n_{101} (P_{101})$	$n_{100} (P_{100})$	$n_{10+} (P_{10+})$
Summe	$n_{1+1} (P_{1+1})$	$n_{1+0} (P_{1+0})$	$n_{1++} (P_{1++})$

Z = 0			
	$D_2 = 1$	$D_2 = 0$	Summe
$D_1 = 1$	$n_{011} (P_{011})$	$n_{010} (P_{010})$	$n_{01+} (P_{01+})$
$D_1 = 0$	$n_{001} (P_{001})$	$n_{000} (P_{000})$	$n_{00+} (P_{00+})$
Summe	$n_{0+1} (P_{0+1})$	$n_{0+0} (P_{0+0})$	$n_{0++} (P_{0++})$

Tabelle 6.2: Symmetrietabellen für den Vergleich zweier diagnostischer Verfahren D_1 und D_2 getrennt nach $Z = 1$ und $Z = 0$. Mit n sind die absoluten Häufigkeiten und mit P die zugehörigen Wahrscheinlichkeiten bezeichnet.

Die Sensitivitäten der beiden Verfahren D_1 und D_2 sind mit dieser Notation gegeben durch:

$$\text{Se}(D_1) = P(D_1 = 1|Z = 1) = \frac{P_{11+}}{P_{1++}} = \frac{P_{11+}}{p} = P_{1+|1},$$

$$\text{Se}(D_2) = P(D_2 = 1|Z = 1) = \frac{P_{1+1}}{P_{1++}} = \frac{P_{1+1}}{p} = P_{+1|1}.$$

Für die Spezifitäten ergibt sich entsprechend:

$$\text{Sp}(D_1) = P(D_1 = 0|Z = 0) = \frac{P_{00+}}{1 - P_{1++}} = \frac{P_{00+}}{1 - p} = P_{0+|0},$$

$$\text{Sp}(D_2) = P(D_2 = 0|Z = 0) = \frac{P_{0+0}}{1 - P_{1++}} = \frac{P_{0+0}}{1 - p} = P_{+0|0}.$$

Zugehörige Schätzer, die empirischen Versionen von $\text{Se}(D_1)$, $\text{Se}(D_2)$, $\text{Sp}(D_1)$ und $\text{Sp}(D_2)$, lauten

$$\hat{\text{Se}}(D_1) = \frac{n_{111} + n_{110}}{n_{1++}}, \quad \hat{\text{Se}}(D_2) = \frac{n_{111} + n_{101}}{n_{1++}},$$

$$\hat{\text{Sp}}(D_1) = \frac{n_{001} + n_{000}}{n_{0++}} \quad \text{und} \quad \hat{\text{Sp}}(D_2) = \frac{n_{010} + n_{000}}{n_{0++}}.$$

6.1.1 Simultaner Vergleich von Sensitivitäten und Spezifitäten

Da zur Validierung des neuen Alarm-Algorithmus gleichzeitig die Nicht-Unterlegenheit bezüglich der Sensitivität und die Überlegenheit bezüglich der Spezifität gezeigt werden

soll, werden im Folgenden Methoden zum simultanen Vergleich der beiden Gütemaße vorgestellt.

Vergleich mittels Intersection-Union-Test

Eine Möglichkeit Sensitivitäten und Spezifitäten simultan zu vergleichen, bietet das Prinzip des Intersection-Union-Tests (Berger und Hsu, 1996), da sich die Nullhypothese als Vereinigung (Union) zweier Einzel-Nullhypothesen und die Alternative als Schnitt (Intersection) zweier Einzel-Alternativen formulieren lässt:

Das Nicht-Unterlegenheits-Testproblem in Bezug auf die Sensitivität (Se) ist gegeben durch

$$H_0^{\text{Se}} : \text{Se}(D_1) - \text{Se}(D_2) \geq \delta \quad \text{vs.} \quad H_1^{\text{Se}} : \text{Se}(D_1) - \text{Se}(D_2) < \delta, \quad (6.4)$$

wobei δ die gewählte Nicht-Unterlegenheitsschranke darstellt. In der vorliegenden Studie ist $\delta = 0.05$.

Das Überlegenheits-Testproblem in Bezug auf die Spezifität (Sp) hat die Form:

$$H_0^{\text{Sp}} : \text{Sp}(D_1) - \text{Sp}(D_2) \geq 0 \quad \text{vs.} \quad H_1^{\text{Sp}} : \text{Sp}(D_1) - \text{Sp}(D_2) < 0. \quad (6.5)$$

Da Nicht-Unterlegenheit bezüglich der Sensitivität und Überlegenheit bezüglich der Spezifität *gleichzeitig* gezeigt werden sollen, lautet das Intersection-Union-Testproblem (Lu, Jin und Genant, 2003)

$$H_0 : H_0^{\text{Se}} \cup H_0^{\text{Sp}} \quad \text{vs.} \quad H_1 : H_1^{\text{Se}} \cap H_1^{\text{Sp}}. \quad (6.6)$$

Ein Intersection-Union-Test zu diesem Testproblem lehnt H_0 genau dann zum Niveau α ab, wenn H_0^{Se} und H_0^{Sp} jeweils einzeln zum Niveau α abgelehnt werden können.

Zur Überprüfung von H_0^{Se} gegen H_1^{Se} kann eine Score-Teststatistik konstruiert werden (siehe Nam, 1997 bzw. Tango, 1998). Diese Teststatistik schneidet in Vergleichen mit anderen in der Literatur vorgeschlagenen Teststatistiken hinsichtlich der Einhaltung des nominal vorgegebenen Signifikanzniveaus am besten ab (siehe z.B. Liu et al., 2002, Tang, 2003). Die Score-Teststatistik ist gegeben durch

$$X_S(\delta) = \frac{n_{110} - n_{101} - n_{1++}\delta}{\sqrt{n_{1++}(2\tilde{p}_{01|1} + \delta - \delta^2)}}, \quad (6.7)$$

wobei $\tilde{p}_{01|1} = ((B^2 - 4AC)^{1/2} - B)/(2A)$, $A = 2n_{1++}$, $B = (2n_{1++} + n_{101} - n_{110})\delta - (n_{110} + n_{101})$ und $C = -n_{101}\delta(1 - \delta)$.

$X_S(\delta)$ ist für $\text{Se}(D_1) - \text{Se}(D_2) = \delta$ bei Unabhängigkeit der Beobachtungspaare asymptotisch standardnormalverteilt. H_0^{Se} kann somit zum Niveau α abgelehnt werden, falls $X_S(\delta) < u_\alpha$, wobei u_α das α -Quantil der Standardnormalverteilung bezeichnet.

Der vorgestellte Test ist eine Verallgemeinerung des McNemar-Tests für den Fall eines Nicht-Unterlegenheitsproblems, da für $\delta = 0$ in (6.7) die unquadrierte McNemar-Teststatistik (siehe Gleichung (6.11) unten) resultiert. Der Test ist auch anwendbar, wenn eine der zu vergleichenden Sensitivitäten maximal ist (siehe Tango, 1998). Dies ist insbesondere für die vorliegende klinische Studie bedeutsam, da das derzeit eingesetzte Schwellwertalarmsystem aufgrund des Studiendesigns eine Sensitivität von 100% aufweist. Berücksichtigt man diese Tatsache, kann das Nicht-Unterlegenheits-Testproblem aus (6.4) allerdings in diesem speziellen Fall auch umformuliert werden zu

$$H_0^{\text{Se}} : \text{Se}(D) \leq \text{Se}_0 \quad \text{vs.} \quad H_1^{\text{Se}} : \text{Se}(D) > \text{Se}_0, \quad \text{mit} \quad \text{Se}_0 = 0.95. \quad (6.8)$$

Das Testproblem reduziert sich somit auf einen Vergleich der Sensitivität des diagnostischen Verfahrens $D = D_2$ mit einem fest vorgegebenen Wert. Dieses Testproblem kann z.B. mit Hilfe der Wald-Teststatistik

$$X_W = \frac{\hat{\text{Se}}(D) - \text{Se}_0}{\sqrt{\text{Var}(\hat{\text{Se}}(D))}} = \frac{\hat{\text{Se}}(D) - \text{Se}_0}{\sqrt{\hat{\text{Se}}(D)(1 - \hat{\text{Se}}(D))/n_{1+}}} \quad (6.9)$$

oder der Score-Teststatistik

$$X_S = \frac{\hat{\text{Se}}(D) - \text{Se}_0}{\sqrt{\text{Se}_0(1 - \text{Se}_0)/n_{1+}}} \quad (6.10)$$

behandelt werden (siehe Agresti, 2002, S. 14). Dabei ist n_{1+} die Anzahl der Beobachtungen, für die $Z = 1$ gilt. X_W und X_S sind für $\text{Se}(D) = \text{Se}_0$ jeweils asymptotisch standardnormalverteilt, so dass H_0^{Se} zum Niveau α abgelehnt werden kann, falls der Wert der jeweiligen Teststatistik größer als $u_{1-\alpha}$ ist.

Zur Überprüfung des Überlegenheits-Testproblems für die Spezifitäten kann der Test von McNemar – ein approximativer zweiseitiger Test auf Symmetrie für (2×2) -Tafeln – verwendet werden (siehe McNemar, 1947). Zur Überprüfung von einseitigen Hypothesen, wie in diesem Fall, kann die zugehörige nicht quadrierte Teststatistik eingesetzt werden. Für das Testproblem H_0^{Sp} gegen H_1^{Sp} ist die Teststatistik gegeben durch

$$X_M = \frac{n_{001} - n_{010}}{\sqrt{n_{001} + n_{010}}}. \quad (6.11)$$

X_M ist für $\text{Sp}(D_1) - \text{Sp}(D_2) = 0$ bei Unabhängigkeit der Beobachtungspaare approximativ standardnormalverteilt. H_0^{Sp} wird somit zum Signifikanzniveau α abgelehnt, falls $X_M < u_\alpha$.

Wird mit Hilfe der vorgestellten Teststatistiken sowohl die Nicht-Unterlegenheit bzgl. der Sensitivität als auch die Überlegenheit bzgl. der Spezifität gezeigt, kann gemäß

des Intersection-Union-Testprinzips auch die Nullhypothese $H_0 : H_0^{\text{Se}} \cup H_0^{\text{Sp}}$ abgelehnt werden.

Zusätzlich können Konfidenzintervalle für die Differenz der Sensitivitäten bzw. Spezifitäten konstruiert werden. Dabei sind Intervalle basierend auf der Invertierung des Score-Tests (Tango, 1998, Agresti, 2002, S. 411) hinsichtlich der Einhaltung der nominal vorgegebenen Überdeckungswahrscheinlichkeit gut geeignet. Für Sensitivitäten und Spezifitäten selbst können z.B. die von Wilson (1927) vorgeschlagenen Score-Konfidenzintervalle verwendet werden (siehe auch Agresti, 2002, S. 16).

Aus Sensitivität und Spezifität kombinierte Gütemaße

Es gibt verschiedene Möglichkeiten, die Sensitivität und Spezifität eines diagnostischen Verfahrens zu einem Gütemaß zu kombinieren. Anstatt für zwei Diagnoseverfahren jeweils Sensitivitäten und Spezifitäten zu vergleichen, kann ein simultaner Vergleich auch anhand eines solchen kombinierten Gütemaßes erfolgen.

Ein bekanntes Maß ist die sogenannte Genauigkeit, auch Accuracy (Ac) oder Treffsicherheit genannt. Sie gibt die Wahrscheinlichkeit für eine korrekte Diagnose an:

$$\text{Ac}(D) = P(D = 1 \cap Z = 1) + P(D = 0 \cap Z = 0)$$

und kann geschätzt werden durch die Summe der Anteile der richtig positiven und richtig negativen Entscheidungen in der Stichprobe (siehe Zhou, Obuchowski und McClish, 2002, S. 22). Die Genauigkeit ist ein gewichtetes Mittel aus Sensitivität und Spezifität

$$\text{Ac}(D) = p\text{Se}(D) + (1 - p)\text{Sp}(D),$$

wobei die Gewichte auf der Prävalenz p des zu diagnostizierenden Zustandes basieren. Die Genauigkeiten zweier diagnostischer Verfahren können mittels Signifikanztests verglichen werden (siehe z.B. Lu, Jin und Genant, 2003). Für die vorliegende klinische Studie ist ein Vergleich von Alarmsystemen anhand der Accuracy allerdings ungeeignet. Zwei Diagnoseverfahren können dieselbe Genauigkeit aufweisen, aber gänzlich unterschiedliche Sensitivitäten und Spezifitäten besitzen. Im Fall des Vergleichs zweier Alarm-Algorithmen ist es aber insbesondere in Bezug auf die Sensitivität wichtig, dass kein zu großer Unterschied zwischen dem alten und dem neu entwickelten Alarmsystem besteht. Somit ist in diesem Fall ein Gütemaß, bei dem ein großer Sensitivitätsverlust möglicherweise durch einen großen Spezifitätsgewinn „ausgeglichen“ wird inadäquat. Chen, Hsueh und Liu (2003) und Newcombe (2001) schlagen vor, nicht nur die Prävalenz, sondern auch die relative Wichtigkeit der Sensitivität und Spezifität bzw. die Kosten für falsch positive und falsch negative Entscheidungen in die Gewichtung eingehen zu lassen. Eine getrennte Betrachtung von Sensitivität und Spezifität erscheint

für die Validierung von Alarmsystemen jedoch sinnvoller als eine Vorgehensweise, die von der nicht trivialen Spezifizierung derartiger Gewichte abhängt.

Weitere gebräuchliche und aus Sensitivität und Spezifität zusammengesetzte Gütemaße, wie z.B. der Youden-Index $YI = Se(D) + Sp(D) - 1$ oder das Odds-Ratio $OR = \frac{Se(D)/(1-Se(D))}{(1-Sp(D))/Sp(D)}$ (siehe Zhou, Obuchowski und McClish, 2002, S. 23), sind zwar nicht prävalenzabhängig, weisen aber wie die Genauigkeit den Nachteil auf, dass falsch positive und falsch negative Entscheidungen gleichwertig behandelt werden und sind somit für die Behandlung der interessierenden Fragestellung ungeeignet.

6.1.2 Berücksichtigung von Abhängigkeiten

In Abschnitt 6.1.1 wurden verschiedene Signifikanztests vorgestellt, die alle auf der wichtigen Annahme der Unabhängigkeit der Beobachtungen bzw. Beobachtungspaare basieren. Für die Tests zum Vergleich der Sensitivität eines Diagnoseverfahrens mit einem festen Wert (Teststatistiken X_W (6.9) und X_S (6.10)) wird die Unabhängigkeit der Entscheidungen des diagnostischen Verfahrens vorausgesetzt. Die Signifikanztests zum Vergleich von Sensitivitäten bzw. Spezifitäten zweier Diagnoseverfahren (Teststatistiken $X_S(\delta)$ (6.7) und X_M (6.11)) fordern die Unabhängigkeit der Entscheidungspaare. Im Falle der Bewertung und des Vergleichs von Alarmsystemen ist diese Annahme möglicherweise nicht gerechtfertigt, da hier viele Beobachtungen bzw. Beobachtungspaare (d.h. Paare von Alarmsystem-Entscheidungen für dieselbe Situation) für denselben Patienten vorkommen und zudem möglicherweise noch zeitliche Abhängigkeiten vorliegen. So weisen bei alarmrelevanten Zuständen für denselben Patienten die zugehörigen Zeitreihen möglicherweise ein ähnliches Verhalten auf, was auch dazu führen kann, dass ein Alarm-Algorithmus in den betreffenden Situationen die gleiche Entscheidung trifft. Auch bei diagnostischen Verfahren können Abhängigkeiten vorkommen. Beispiele sind Studien aus der Augenheilkunde, in denen beide Augen eines Patienten untersucht werden oder Radiologie-Studien, bei denen unterschiedliche Stellen im Körper desselben Menschen mittels diagnostischer Verfahren beurteilt werden. Verwendet man in solchen Fällen die in Abschnitt 6.1.1 angegebenen Teststatistiken, können fälschlicherweise signifikante Testentscheidungen resultieren. Die Ursache hierfür liegt in einer Unterschätzung der in die Teststatistik eingehenden Varianz, wodurch der Absolutwert der Teststatistik größer wird.

In der Literatur zum statistischen Vergleich von Diagnoseverfahren finden sich verschiedene Möglichkeiten, mit Abhängigkeiten umzugehen. Auf diese wird in den folgenden Abschnitten näher eingegangen. Man bezeichnet dann Individuen, bei denen die zu ver-

gleichenden diagnostischen Verfahren an mehreren Stellen oder wiederholt angewendet werden, üblicherweise als *Cluster*. Im Folgenden bezeichne G die Anzahl der vorliegenden Cluster. Lösungsvorschläge, die nachfolgend für Sensitivitäten erläutert werden, gelten analog auch für Spezifitäten und umgekehrt. Wird ein Sensitivitäts-Testproblem betrachtet, bezeichne n_g die Anzahl der Beobachtungen bzw. Beobachtungspaare in Cluster $g = 1, \dots, G$, für die $Z = 1$ gilt. Im Falle eines Spezifitäts-Testproblems gibt n_g die Anzahl Beobachtungen bzw. Beobachtungspaare in Cluster g an, für die $Z = 0$ ist.

Ansätze für die Behandlung von Testproblem (6.8)

Für das Testproblem zum Vergleich der Sensitivität eines Diagnoseverfahrens D mit einem fest vorgegebenen Wert in (6.8) existieren Vorschläge zur alternativen Schätzung der Varianz $\text{Var}(\hat{\text{Se}}(D))$.

Donner und Klar (1993) adjustieren die Varianz durch Multiplikation mit dem „Varianz-Aufblähungsfaktor“ (inflation factor)

$$K = 1 + \left(\frac{\sum_{g=1}^G n_g^2}{n_{1+}} - 1 \right) \hat{\rho},$$

der auf einem Varianzanalyse-Schätzer für den Intraclass-Korrelationskoeffizienten ρ basiert (siehe Fleiss, 1981, S. 226):

$$\hat{\rho} = \frac{\text{BMS} - \text{WMS}}{\text{BMS} + (n_0 - 1)\text{WMS}}. \quad (6.12)$$

Dabei ist

$$\text{BMS} = \sum_{g=1}^G n_g (\hat{\text{Se}}(D)_g - \hat{\text{Se}}(D))^2 / (G - 1)$$

der mittlere quadratische Fehler zwischen den Clustern,

$$\text{WMS} = \sum_{g=1}^G n_g \hat{\text{Se}}(D)_g (1 - \hat{\text{Se}}(D)_g) / (n_{1+} - G)$$

der mittlere quadratische Fehler innerhalb der Cluster, $\hat{\text{Se}}(D)_g$ die geschätzte Sensitivität für Cluster g ,

$$n_0 = \left(n_{1+} - \sum_{g=1}^G \frac{n_g^2}{n_{1+}} \right) / (G - 1)$$

und $n_{1+} = \sum_{g=1}^G n_g$. Die für $\text{Se}(D) = \text{Se}_0$ asymptotisch standardnormalverteilte adjustierte Wald-Teststatistik lautet dann:

$$X_{W'} = \frac{\hat{\text{Se}}(D) - \text{Se}_0}{\sqrt{K \cdot \hat{\text{Se}}(D)(1 - \hat{\text{Se}}(D))/n_{1+}}}. \quad (6.13)$$

Da für die Sensitivität im Falle von G Clustern gilt

$$\hat{Se}(D) = \frac{\sum_{g=1}^G n_g \hat{Se}(D)_g}{\sum_{g=1}^G n_g}$$

und sie somit als Verhältnis von zwei Stichprobenmitteln ausgedrückt werden kann, ergibt sich aus der Stichprobentheorie ein weiterer Schätzer für $\text{Var}(\hat{Se}(D))$, der die clusterinduzierte Abhängigkeitsstruktur berücksichtigt:

$$V_C = \frac{G}{G-1} \sum_{g=1}^G \left(\frac{n_g}{n_{1+}} \right)^2 (\hat{Se}(D)_g - \hat{Se}(D))^2 \quad (6.14)$$

(siehe Cochran, 1977, S. 31 und Rao und Scott, 1992). Hier lautet die entsprechend adjustierte Wald-Teststatistik (für $Se(D) = Se_0$ approximativ standardnormalverteilt):

$$X_{W''} = \frac{\hat{Se}(D) - Se_0}{\sqrt{V_C}}. \quad (6.15)$$

Bei den beiden bisher vorgestellten Methoden erhalten alle Untersuchungseinheiten dasselbe Gewicht, was einer Gewichtung der Cluster proportional zu ihrer Größe entspricht. Lee und Dubin (1994) wählen einen anderen Ansatz und weisen allen Clustern – unabhängig von ihrer Größe – dasselbe Gewicht zu. Die Sensitivität bestimmt sich dann als Durchschnitt der Sensitivitäten der G Cluster. Die Autoren geben auch einen auf dieser Gewichtung basierenden Varianzschätzer an. Ahn (1997) zeigt in einer Simulationsstudie, dass letztere Gewichtungsmethode bei Intracluster-Korrelationen über 0.6 zu bevorzugen ist. Jung und Ahn (2000) schlagen eine alternative Gewichtung vor, die sowohl das Ausmaß der Intraclass-Korrelation als auch die Clustergröße berücksichtigt. Die Gewichte werden durch Varianz-Minimierung bestimmt und hängen vom Intraclass-Korrelationskoeffizienten (6.12) ab.

Ansätze für die Behandlung der Testprobleme (6.4) und (6.5)

Auch für Testprobleme zum Vergleich zweier Sensitivitäten bzw. Spezifitäten (siehe (6.4) und (6.5)) existieren Ansätze für die Berücksichtigung von Abhängigkeiten innerhalb der Cluster, wobei einige den bereits oben beschriebenen Methoden ähneln. Für die Behandlung von Überlegenheits-Testproblemen werden im Folgenden vier Methoden vorgestellt. Zum Vorgehen bei Nicht-Unterlegenheits-Fragestellungen findet sich in der Literatur nur ein Ansatz.

Eliasziw und Donner (1991) adjustieren die McNemar-Teststatistik mit einem Varianz-Aufblähungsfaktor:

$$X_E = \frac{X_M}{\sqrt{1 + (n_c - 1)\hat{\rho}}}. \quad (6.16)$$

Hierbei ist

$$\hat{\rho} = \frac{\text{BMS} - \text{WMS}}{\text{BMS} + (S_0 - 1)\text{WMS}}$$

ein konsistenter Schätzer für den Intraclass-Korrelationskoeffizienten der diskordanten Paare, d.h. der Paare bei denen die Diagnoseverfahren unterschiedliche Ergebnisse aufweisen, und

$$n_c = S_0 + G_d(\bar{S} - S_0)$$

ist die adjustierte mittlere Anzahl diskordanter Paare in der Stichprobe mit

BMS: Mittlerer quadratischer Fehler zwischen den Clustern,

WMS: Mittlerer quadratischer Fehler innerhalb der Cluster,

$$S_0 = \bar{S} - \frac{\sum_{g=1}^G (S_g - \bar{S})^2 - (G - G_d)\bar{S}^2}{G_d(G_d - 1)\bar{S}},$$

$$\bar{S} = \frac{1}{G_d} \sum_{g=1}^G S_g,$$

S_g : Anzahl diskordanter Paare in Cluster g ,

G_d : Anzahl der Cluster mit wenigstens einem diskordanten Paar.

X_E ist bei Gleichheit der Spezifitäten asymptotisch standardnormalverteilt. Die Adjustierung der McNemar-Teststatistik basiert auf der Schätzung der Korrelation ρ unter den diskordanten Paaren. Wenn die Anzahl der diskordanten Paare klein ist, kann diese Schätzung unzuverlässig werden. Eliasziw und Donner (1991) schlagen daher einen alternativen Schätzer, der auf allen Beobachtungspaaren basiert und eine zweite adjustierte McNemar-Teststatistik $X_{E'}$ vor. Die Autoren empfehlen die Anwendung der Methode, wenn die Korrelation in den Clustern nicht abhängig von der Clustergröße ist.

Obuchowski (1998) nutzt die Darstellungsweise für die Spezifität als Verhältnis zweier Stichprobenmittel zur Berechnung von Varianz

$$\text{Var}(\hat{\text{Sp}}(D_i))_{\bar{\text{Sp}}} = \frac{G}{G-1} \sum_{g=1}^G \left(\frac{n_g}{n_{0++}} \right)^2 (\hat{\text{Sp}}(D_i)_g - \bar{\text{Sp}})^2, \quad i = 1, 2$$

und Kovarianz

$$\text{Cov}(\hat{\text{Sp}}(D_1), \hat{\text{Sp}}(D_2))_{\bar{\text{Sp}}} = \frac{G}{G-1} \sum_{g=1}^G \left(\frac{n_g}{n_{0++}} \right)^2 (\hat{\text{Sp}}(D_1)_g - \bar{\text{Sp}})(\hat{\text{Sp}}(D_2)_g - \bar{\text{Sp}})$$

gemäß Cochran (1977, S. 31) und Rao und Scott (1992). Dabei ist $\bar{\text{Sp}} = (\hat{\text{Sp}}(D_1) + \hat{\text{Sp}}(D_2))/2$ die gepoolte Spezifität. Die Teststatistik wird dann zu

$$X_O = \frac{\hat{\text{Sp}}(D_1) - \hat{\text{Sp}}(D_2)}{\sqrt{\text{Var}(\hat{\text{Sp}}(D_1) - \hat{\text{Sp}}(D_2))_{\bar{\text{Sp}}}}}, \quad (6.17)$$

wobei

$$\text{Var}(\hat{\text{Sp}}(D_1) - \hat{\text{Sp}}(D_2))_{\overline{\text{Sp}}} = \text{Var}(\hat{\text{Sp}}(D_1))_{\overline{\text{Sp}}} + \text{Var}(\hat{\text{Sp}}(D_2))_{\overline{\text{Sp}}} - 2\text{Cov}(\hat{\text{Sp}}(D_1), \hat{\text{Sp}}(D_2))_{\overline{\text{Sp}}}.$$

X_O ist für $\text{Sp}(D_1) - \text{Sp}(D_2) = 0$ asymptotisch standardnormalverteilt. In einer Simulationsstudie, bei der drei verschiedene Korrelationsstrukturen betrachtet werden, vergleicht Obuchowski diese Teststatistik mit den beiden von Eliasziw und Donner vorgeschlagenen Testverfahren. Hier zeigt sich, dass X_O bevorzugt werden sollte, wenn die Clustergrößen stark variieren oder wenn nicht angenommen werden kann, dass die Korrelation in den Clustern immer gleich ist.

Durkalski et al. (2003a) schlagen als Teststatistik

$$X_R = \frac{\sum_{g=1}^G (\hat{\text{Sp}}(D_1)_g - \hat{\text{Sp}}(D_2)_g)}{\sqrt{\sum_{g=1}^G (\hat{\text{Sp}}(D_1)_g - \hat{\text{Sp}}(D_2)_g)^2}} \quad (6.18)$$

vor. X_R ist bei Gleichheit der Spezifitäten approximativ standardnormalverteilt. Die Autoren vergleichen ihren Test mit den Tests basierend auf den oben beschriebenen Teststatistiken $X_{E'}$ und X_O . Simulationen zeigen, dass X_R hinsichtlich der Rate für den Fehler 1. Art und der Power ähnlich wie X_O abschneidet, wenn sich die Clustergrößen nicht um mehr als fünf Einheiten unterscheiden. Dagegen sollte X_O verwendet werden, wenn sich die Clustergrößen stärker unterscheiden und wenn die Spezifitäten in den einzelnen Clustern stark variieren.

Durkalski et al. (2003b) erweitern ihre Methode auf den Fall, dass ein Nicht-Unterlegenheits-Testproblem gelöst werden soll. Die Teststatistik zur Behandlung des Testproblems in (6.4) lautet:

$$X_R(\delta) = \frac{\sum_{g=1}^G (\hat{\text{Se}}(D_1)_g - \hat{\text{Se}}(D_2)_g - \delta)}{\sqrt{\sum_{g=1}^G (\hat{\text{Se}}(D_1)_g - \hat{\text{Se}}(D_2)_g - \delta)^2}}. \quad (6.19)$$

$X_R(\delta)$ ist bei Gleichheit der Spezifitäten für großes G standardnormalverteilt.

Alternativ zu obigen Verfahren können auch Marginale Regressionsmodelle (Liang und Zeger, 1986) einen Lösungsansatz bieten. Leisenring, Pepe und Longton (1997) verwenden diese zur Modellierung von Sensitivität und Spezifität. Sei im Folgenden Y_{gh} das h -te Diagnose-Ergebnis für Cluster g , $g = 1, \dots, G$, $h = 1, \dots, m_g$, wobei unterschiedliche diagnostische Verfahren verwendet werden. D.h. manche der m_g Diagnose-Ergebnisse können sich auf dieselbe Untersuchungseinheit in Cluster g beziehen und Entscheidungen verschiedener Diagnoseverfahren für diese Untersuchungseinheit angeben. Mit Z_{gh} sei der Krankheitsstatus und mit X_{gh} Kovariablen bezeichnet. Dann ist ein allgemeines

Regressionsmodell für die Sensitivität gegeben durch

$$P(Y_{gh} = 1 | Z_{gh} = 1, X_{gh}) = g_Z(\beta_Z, X_{gh}).$$

Ein Regressionsmodell für die falsch positiv Rate (1-Spezifität) ist dementsprechend:

$$P(Y_{gh} = 1 | Z_{gh} = 0, X_{gh}) = g_{\bar{Z}}(\beta_{\bar{Z}}, X_{gh}).$$

Hierbei sind β_Z und $\beta_{\bar{Z}}$ Parametervektoren und g_Z und $g_{\bar{Z}}$ bekannte Regressionsfunktionen. Es wird der Spezialfall betrachtet, dass zwei Diagnoseverfahren verglichen werden sollen. Sei X eine binäre Variable mit $X = 0$, falls das diagnostische Verfahren D_1 und $X = 1$, falls Verfahren D_2 verwendet wird. Werden für jede Untersuchungseinheit in Cluster g beide Diagnoseverfahren durchgeführt, bezieht sich die eine Hälfte der m_g Diagnosen auf Verfahren D_1 und die andere Hälfte auf Verfahren D_2 . Dann sind mögliche Modelle für Sensitivität und (1-Spezifität) gegeben durch:

$$\text{logit } P(Y_{gh} = 1 | Z_{gh} = 1, X_{gh}) = \beta_Z^0 + \beta_Z^1 X_{gh} \quad (6.20)$$

und

$$\text{logit } P(Y_{gh} = 1 | Z_{gh} = 0, X_{gh}) = \beta_{\bar{Z}}^0 + \beta_{\bar{Z}}^1 X_{gh}. \quad (6.21)$$

Im Modell (6.20) ist ein Test für $H_0 : \beta_Z^1 = 0$ äquivalent zu einem Test für H_0 : Sensitivität(D_1) = Sensitivität(D_2). Entsprechend gilt für (6.21), dass $H_0 : \beta_{\bar{Z}}^1 = 0$ äquivalent ist zu H_0 : Spezifität(D_1) = Spezifität(D_2).

In dem einfachen Spezialfall wurde als Kovariable nur eine Variable eingeführt, die Auskunft über das angewandte Diagnoseverfahren gibt. Für manche Untersuchungen ist es auch denkbar, weitere Faktoren zu analysieren, die möglicherweise einen Einfluss auf die Sensitivität und Spezifität haben und diese somit als Kovariablen in das Modell aufzunehmen. Die so aufgestellten Regressionsmodelle liefern Informationen über Sensitivität und Spezifität der beiden diagnostischen Verfahren unter verschiedenen Umständen. Dies kann Anhaltspunkte dafür liefern unter welchen Bedingungen die jeweiligen Verfahren optimalerweise durchgeführt werden sollten.

Marginale Regressionsmodelle können mit Hilfe von Generalized Estimating Equations (GEE) angepasst werden (Liang und Zeger, 1986). Für die bei dieser Methode zu spezifizierende sogenannte „Arbeits-Varianz-Kovarianzmatrix“ empfehlen Leisenring, Pepe und Longton (1997) die Verwendung einer Diagonal-Matrix, bei der auf der Hauptdiagonale die Varianzen der Y_{gh} stehen, so dass zwischen den Beobachtungen Unabhängigkeit angenommen wird. Für diese Wahl können die Modelle für Sensitivität und Spezifität getrennt angepasst werden. Auch bei Fehlspezifikation der Kovarianzstruktur sind die Parameterschätzer konsistent, obwohl eine korrekte Wahl zu einer

höheren Effizienz führt. Für die Schätzung der Kovarianzmatrix für die Parameter muss die durch die Clusterstruktur gegebene Abhängigkeit allerdings berücksichtigt werden. Die Schätzung erfolgt dann mittels des sogenannten „Sandwich-Varianzschätzers“ (Liang und Zeger, 1986). Dieser kann zur Konstruktion einer Wald-Teststatistik verwendet werden. Außerdem kann ein generalisierter Score-Test hergeleitet werden, der eine Verallgemeinerung des McNemar-Tests darstellt.

6.2 Übertragung der Methodik auf den Vergleich von Alarmsystemen

Intensivmedizinische Monitoringsysteme weisen im Vergleich zu den im vorangegangenen Abschnitt betrachteten Diagnoseverfahren die Besonderheit auf, dass kontinuierlich geprüft wird, ob ein alarmierungswürdiger Zustand vorliegt oder nicht. Bei einfachen diagnostischen Verfahren gibt es dagegen klar definierte Untersuchungseinheiten, etwa Patienten bzw. unterschiedliche Körperstellen eines Patienten, auf die das Diagnoseverfahren angewandt wird. Die Aufbereitung der erhobenen Daten mit Kontingenz- oder Symmetrietabellen und die Schätzung von Gütemaßen und deren Vergleich ist dann problemlos möglich. Im Falle von Alarmsystemen müssen Untersuchungseinheiten erst geeignet definiert werden – beispielsweise als Zeitintervalle. In den folgenden Abschnitten soll die im vorangegangenen Abschnitt vorgestellte Methodik auf Alarmsysteme übertragen werden.

6.2.1 Bestimmung der Sensitivität

Die Sensitivität bezüglich der Erkennung alarmrelevanter Zustände kann für ein Alarmsystem geschätzt werden durch das Verhältnis von erkannten alarmrelevanten Situationen und allen aufgetretenen alarmrelevanten Situationen. Wie bereits erwähnt, ist die Sensitivität des derzeit verwendeten Alarmsystems in dieser klinischen Studie gleich 100%, da sich die Annotationen ausschließlich auf die Alarmzeitpunkte des derzeit verwendeten Alarmsystems beziehen und somit während eines alarmrelevanten Zustandes immer auch ein Alarm des alten Alarmsystems vorkommt.

Für das neue System muss eine gesonderte Vorgehensweise zur Bestimmung der Sensitivität festgelegt werden. Insbesondere muss definiert werden, wann ein alarmrelevanter Zustand als vom neuen System *erkannt* angesehen wird. Da sich die An-

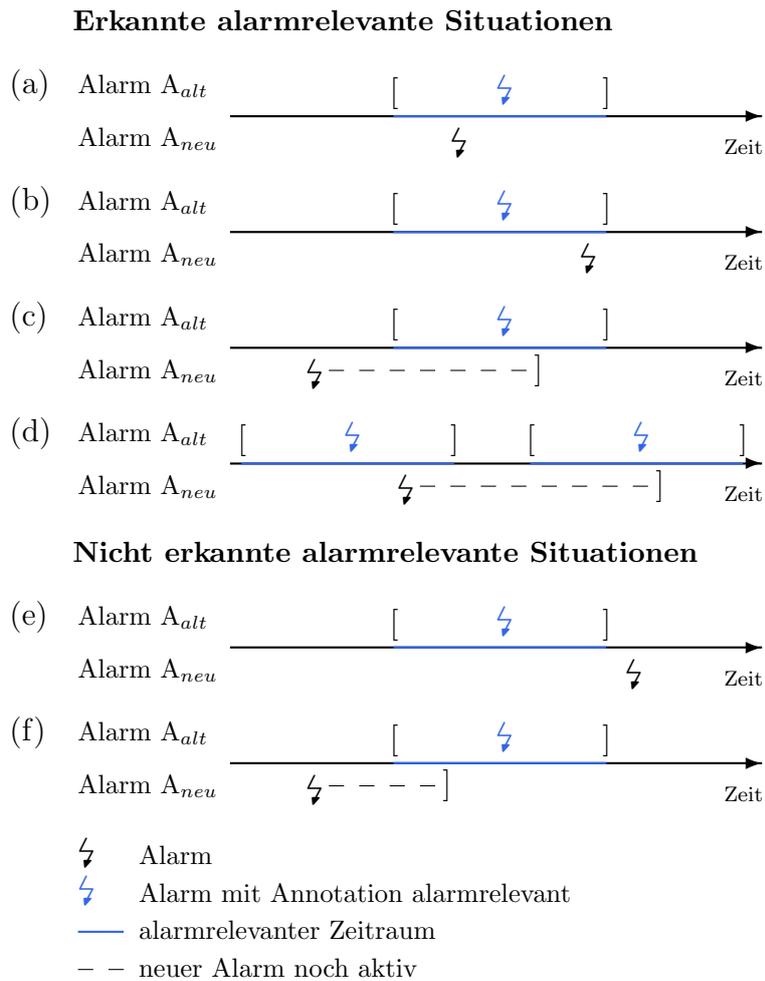


Abbildung 6.1: Beispiele für erkannte und nicht erkannte alarmrelevante Situationen. A_{alt} bezeichnet das derzeit eingesetzte Alarmsystem und A_{neu} bezeichnet den neu entwickelten Alarm-Algorithmus.

notationen nur auf einzelne Zeitpunkte beziehen und ein neuer Alarm-Algorithmus im Allgemeinen nicht in genau derselben Sekunde alarmieren wird, liegt die Aufgabe in der Festlegung alarmrelevanter Zeiträume ausgehend von den Alarmen des Infinity Monitoringsystems[®] mit der Bewertung alarmrelevant. Um diese sind Intervalle zu legen, die so gewählt sein müssen, dass ein neuer Alarm – unabhängig davon zu welchem Zeitpunkt er innerhalb des Intervalls auftritt – den alarmrelevanten Zustand rechtzeitig erkennt. In dieser Studie werden zur Konstruktion der alarmrelevanten Zeiträume Intervalle symmetrisch um jeden alarmrelevant bewerteten Alarm gelegt, wobei sich die Intervalllänge nach dem alarmlösenden Vitalparameter richtet. Aus medizinischer Sicht sind für die Erkennung alarmrelevanter Situationen, die den invasiven arteriellen Blutdruck oder die Sauerstoffsättigung betreffen, z.B. Intervalle der Länge ± 60 Sekunden angemessen. Alarmrelevante Zustände aufgrund einer

Veränderung der Herzfrequenz sind dagegen zeitnäher zu erkennen, weshalb für diesen Vitalparameter eine Intervalllänge von ± 30 Sekunden gewählt wird.

Alarmrelevante Situationen, bei denen innerhalb des festgelegten Intervalls kein Alarm des neuen Alarm-Algorithmus auftritt, werden trotzdem als erkannt angesehen, falls ein neuer Alarm zwar vor Beginn des festgelegten Intervalls anfängt, aber zum Auslösungszeitpunkt des alarmrelevant annotierten Schwellwertalarms *noch aktiv* ist.

In Abbildung 6.1 sind Beispiele für erkannte alarmrelevante Situationen schematisch dargestellt. Abbildung 6.1(a) und (b) zeigen Fälle, in denen neue Alarme innerhalb des Intervalls ausgelöst werden. In Abbildung 6.1(c) wird ein neuer Alarm zwar vor Beginn des alarmrelevanten Zeitraums ausgelöst, er dauert aber noch bis zum Auslösungszeitpunkt des Alarms mit Annotation alarmrelevant an. Aus Abbildung 6.1(d) wird deutlich, dass ein neuer Alarm auch mehrere alarmrelevante Situationen erkennen kann. Abbildung 6.1(e) und (f) zeigen Beispiele für nicht erkannte alarmrelevante Situationen. Im ersten Beispiel kommt der neue Alarm zu spät, im zweiten Beispiel ist er nicht lange genug aktiv.

6.2.2 Bestimmung der Spezifität

Die Schätzung der Spezifität eines Alarmsystems erfordert die Bestimmung der Anzahl nicht alarmrelevanter Zustände. Die Spezifität ist dann der Anteil der nicht alarmrelevanten Zustände ohne Alarm an allen nicht alarmrelevanten Zuständen. Hierfür muss zunächst definiert werden, in welchen Zeiträumen der Zustand nicht alarmrelevant ist. In dieser Studie sind dies alle Zeiträume außerhalb der für die Sensitivitätsbestimmung verwendeten Intervalle um einen Alarm mit Bewertung alarmrelevant. Falsch positiv sind dann alle Alarme, die nicht in einen der alarmrelevanten Zeiträume fallen und die während eines alarmrelevant annotierten Alarms nicht mehr aktiv sind. Dieses Prinzip wird in Abbildung 6.2 schematisch verdeutlicht. Die Bezeichnung falsch positiv ist in diesem Zusammenhang im Sinne von „nicht in Bezug zu einem alarmrelevanten Zustand stehend“ gemeint. Zu den falsch positiven Alarmen können daher hier z.B. auch Alarme gehören, die in den Annotationen als hinweisend bewertet wurden. Bei der gewählten Definition kann es auch vorkommen, dass Alarme des Infinity Monitors[®], die in den Annotationen mit hinweisend oder nicht alarmrelevant bewertet wurden, in einen alarmrelevanten Zeitraum fallen und somit nicht zu den falsch positiven Alarmen zählen.

Für die Spezifitätsbestimmung muss der nicht alarmrelevante Zeitraum geeignet aufgeteilt werden, um die Anzahl nicht alarmrelevanter Situationen zu erhalten. Hier liegt

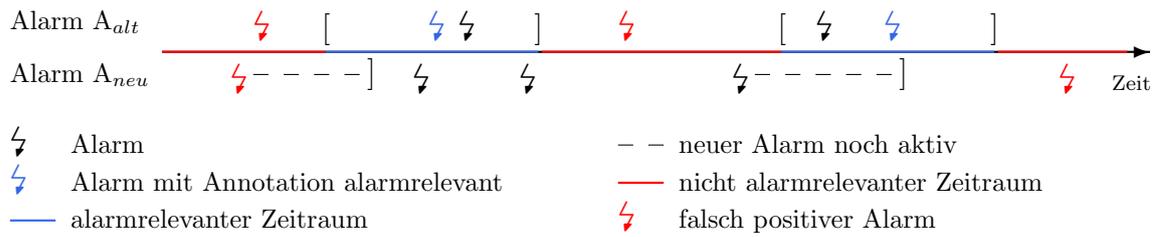


Abbildung 6.2: Alarmrelevante und nicht alarmrelevante Zeiträume sowie falsch positive Alarme

das große prinzipielle Problem der Spezifitätsbestimmung für Alarmsysteme. Die Art und Weise der Aufteilung beeinflusst den Schätzwert für die Spezifität entscheidend. Dies hat zur Folge, dass die Angabe eines Spezifitätswertes für ein Alarmsystem ohne gleichzeitige Nennung der Bestimmungsweise für die Anzahl nicht alarmrelevanter Zeiträume praktisch keine Aussagekraft besitzt. Dies soll anhand des folgenden – extrem gewählten – Beispiels verdeutlicht werden: Hierfür wird die intuitiv überzeugende Aufteilung des nicht alarmrelevanten Zeitraums, nämlich die Aufteilung in gleich große Zeitabschnitte verwendet. Für jeden Zeitabschnitt wird jeweils bestimmt, ob das Alarmsystem einen Alarm auslöst oder nicht. Jeder Abschnitt mit Alarm wird als falsch positive Beobachtung und jeder Abschnitt ohne Alarm als richtig negative Beobachtung gezählt. Betrachtet wird die in Abbildung 6.3 dargestellte Ausgangssituation. In dem betreffenden nicht alarmrelevanten Zeitraum treten fünf unerwünschte Alarme auf. Eine erste Aufteilung in gleich große Abschnitte liefert eine geschätzte Spezifität von 0%, da in jedes der gebildeten Intervalle ein Alarm fällt. Wählt man eine feinere Aufteilung durch Halbierung der Abschnittslänge, vergrößert sich die Spezifität auf 50%, da nun Intervalle ohne Alarm vorkommen. Eine weitere Halbierung der Abschnittslänge liefert gar eine Spezifität von 75%. Dies kann fortgeführt werden, bis jeder Abschnitt der kleinsten Zeiteinheit entspricht, d.h. eine Sekunde lang ist. Für diesen Fall würde man eine Spezifität sehr nahe an 100% erhalten. Extrem große oder kleine Werte für die Abschnittslänge sollten daher vermieden werden. Plausibel erscheinen z.B. Werte

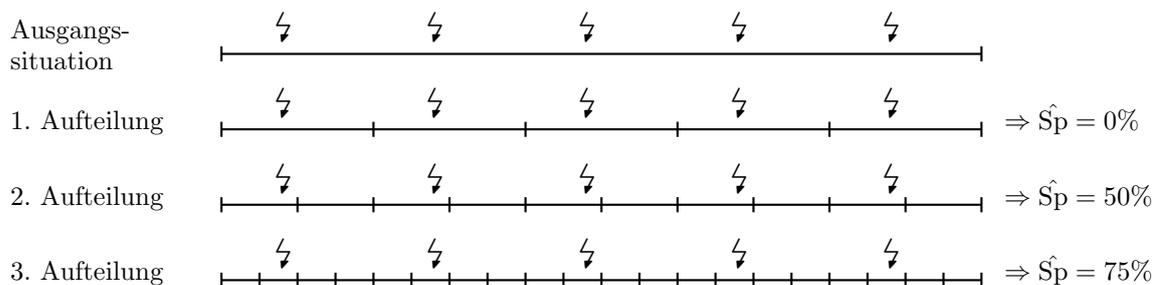


Abbildung 6.3: Extremes Beispiel für die Beeinflussbarkeit der Spezifität

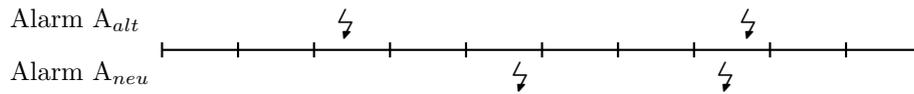


Abbildung 6.4: Aufteilung des nicht alarmrelevanten Zeitraums in gleich große Zeitabschnitte

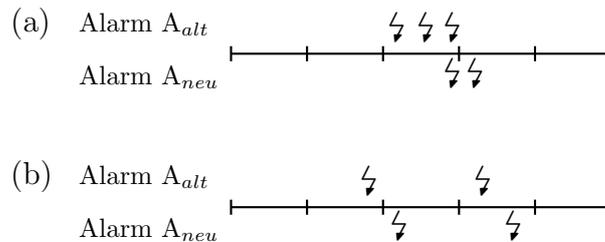


Abbildung 6.5: Probleme bei der Aufteilung des nicht alarmrelevanten Zeitraums in gleich große Zeitabschnitte

zwischen einer Minute und zehn Minuten.

Die Aufteilung in Abschnitte hat beim Vergleich zweier Alarmsysteme den Vorteil, dass jeder Zeitabschnitt eine Untersuchungseinheit bildet, für die jeweils bestimmt werden kann, ob vom derzeitig eingesetzten Alarmsystem und/oder vom neuen Alarm-Algorithmus ein Alarm gegeben wird (siehe Abbildung 6.4). Mittels dieser Informationen lässt sich für die Paare von Alarmsystem-Entscheidungen eine Symmetrietabelle wie in Tabelle 6.2 aufstellen, aus der ersichtlich ist, wie oft die Alarmsysteme eine Situation gleich (konkordant) oder unterschiedlich (diskordant) bewerten. Diese Angaben werden z.B. für die Teststatistik des McNemar-Tests zum Vergleich von Spezifitäten benötigt. Allerdings kann die eigentlich anhand der Spezifität zu untersuchende Frage, ob die Häufigkeit von Fehlalarmen durch einen neuen Alarm-Algorithmus vermindert wird, mittels dieser Vorgehensweise nur sehr bedingt geklärt werden. Schließlich wird für ein Intervall nur erfasst, *ob* ein System einen Alarm auslöst und nicht *wie viele* Alarme ausgelöst werden. Sei hierzu die in Abbildung 6.5(a) dargestellte Situation betrachtet: Obwohl das alte Alarmsystem mehr Fehlalarme als der neue Alarm-Algorithmus erzeugt, würde für das alte Alarmsystem nur eine falsch positive Beobachtung gezählt, während für den neuen Alarm-Algorithmus zwei falsch positive Entscheidungen resultieren würden. Derartige Fälle, in denen viele Alarme für dieselbe Situation in kurzem Abstand zueinander auftreten, sind bei intensivmedizinischen Alarmsystemen keine Seltenheit. Sie treten z.B. oft auf, wenn die Zeitreihe eines Vitalparameters einige Zeit um die Alarmgrenze herum schwankt.

Durch die strenge Aufteilung in Zeitabschnitte kann es außerdem vorkommen, dass zwei Alarme, die von unterschiedlichen Alarmsystemen ausgelöst werden, in einigen Fällen zu konkordanten und in anderen Fällen zu diskordanten Beobachtungen führen, obwohl sie zeitlich gleich weit voneinander entfernt sind (siehe Abbildung 6.5(b)).

Weiterhin kann die vorgegebene Abschnittslänge in manchen Situationen nicht exakt eingehalten werden, zum einen wenn ein Zeitabschnitt zwischen zwei alarmrelevanten Zuständen kürzer als die vorgegebene Abschnittslänge ist und zum anderen wenn bei der Aufteilung eines nicht alarmrelevanten Zeitraums ein Rest verbleibt. Hier müssen Ausnahmeregelungen getroffen werden.

Eine Möglichkeit, die Anzahl falsch positiver Alarme bei der Berechnung der Spezifität zu berücksichtigen, bietet der folgende Ansatz, welcher Ähnlichkeiten zu der von Lawless (1994) gewählten Methode zur Bestimmung der Anzahl richtig negativer Entscheidungen aufweist (siehe Tabelle 5.3):

Zunächst wird ausgehend von der Gesamtdauer der nicht alarmrelevanten Zeiträume festgelegt, wie viele falsch positive Entscheidungen eines Alarmsystems man in diesem Zeitraum schlimmstenfalls insgesamt erwarten würde, z.B. einen Alarm alle fünf Minuten. Diese Anzahl sei mit FP_{max} bezeichnet. Die Anzahl der tatsächlich aufgetretenen falsch positiven Alarme eines Systems, bezeichnet mit FP , wird von dieser „Worst case“-Anzahl abgezogen, um die Anzahl richtig negativer Entscheidungen des Alarm-Algorithmus zu erhalten. Ist die Differenz größer als Null, so ist die Spezifität der Anteil der richtig negativen Entscheidungen an der schlimmstenfalls möglichen Anzahl falsch positiver Entscheidungen. Ist die Differenz kleiner als Null, ist die Spezifität gleich Null. Formal ausgedrückt ergibt sich

$$\hat{Sp} = \begin{cases} \frac{FP_{max} - FP}{FP_{max}}, & FP \leq FP_{max} \\ 0, & FP > FP_{max} \end{cases} \quad (6.22)$$

Bei dieser Definition spielt nur die Anzahl und nicht das zeitliche Auftreten der Fehlalarme eine Rolle. Dadurch ergeben sich Spezifitäten, die das Abschneiden von Alarmsystemen hinsichtlich der Häufigkeit falsch positiver Alarme besser reflektieren als Spezifitäten basierend auf festen Zeitabschnitten. Es ist bei Verwendung von (6.22) nicht möglich, dass bei zwei zu vergleichenden Alarmsystemen das System mit weniger Fehlalarmen eine geringere Spezifität als das System mit mehr Fehlalarmen hat. Allerdings kann der Wert der Spezifität natürlich auch bei diesem Ansatz durch die Wahl der „Worst case“-Anzahl beeinflusst werden. Dies sei an folgendem einfachen Beispiel verdeutlicht:

Die Gesamtlänge der nicht alarmrelevanten Zeiträume betrage 10 Stunden. Das interessierende Alarmsystem löse in dieser Zeit insgesamt $FP = 15$ Fehlalarme aus. Legt man

fest, dass schlimmstenfalls ein falsch positiver Alarm alle 30 Minuten auftreten kann, ergibt sich als „Worst case“-Anzahl $FP_{max} = 10 \text{ Std.}/30 \text{ Min.} = 20$ und die geschätzte Spezifität beträgt $\hat{Sp} = (20 - 15)/20 = 25\%$. Wird allerdings ein potentieller falsch positiver Alarm alle 10 Minuten erwartet, d.h. $FP_{max} = 10 \text{ Std.}/10 \text{ Min.} = 60$, ist die Spezifität deutlich höher, da $\hat{Sp} = (60 - 15)/60 = 75\%$. Um einen plausiblen Wert für die „Worst case“-Anzahl und somit auch für die Spezifität zu erhalten, sollte eine Orientierung an Erfahrungswerten erfolgen. Sind für eine große Anzahl Patienten die aufgetretenen Alarmfrequenzen bekannt, so kann z.B. die maximal beobachtete Alarmfrequenz als Anhaltspunkt für die Bestimmung der schlimmstenfalls in einem Zeitraum erwarteten Alarmanzahl herangezogen werden.

Bei dieser Methode für die Berechnung der Spezifität kann keine Symmetrietabelle wie in Tabelle 6.2 erstellt werden, da durch Nichtberücksichtigung der zeitlichen Reihenfolge keine Aussage über konkordante bzw. diskordante Alarmsystem-Entscheidungen möglich ist. Der McNemar-Test kann für die Überprüfung des Überlegenheits-Testproblems für die Spezifität bei diesem Ansatz somit nicht verwendet werden. Wird jeder Patient als Cluster betrachtet, sind dagegen die Tests mit den Teststatistiken X_O (6.17) bzw. X_R (6.18) sowie die Methode der marginalen Regressionsmodelle anwendbar.

Bei der vorgestellten Definition für falsch positive Alarme erfahren durch Manipulation verursachte Alarme keine gesonderte Behandlung. Wie schon in Abschnitt 5.2 erwähnt, nehmen solche Alarme aber prinzipiell eine Sonderstellung ein, da sie aktiv herbeigeführt werden und somit nicht primär dem Monitorsystem anzulasten sind. Von Interesse ist somit auch die Spezifität eines Alarmsystems ohne Berücksichtigung der manipulierten Alarme. Für die Alarme des Infinity Monitoring Systems[®] ist die Information über eine Manipulation in den Annotationen festgehalten. Da für neu generierte Alarme solche exakten Informationen nicht vorliegen, kann nur über deren zeitliche Nähe zu den herbeigeführten Alarmen des einfachen Schwellwertalarmsystems eine Vermutung darüber getroffen werden, ob sie ebenfalls manipulationsbedingt sind. Da dies wieder die Spezifizierung von Zeitfenstern beinhaltet, was einen zusätzlichen Unsicherheitsfaktor darstellt, wird für die Validierung weiterhin die Spezifität unter Berücksichtigung aller Alarme betrachtet.

6.2.3 Auswahl von Validierungsmethodiken

Nachdem in den vorangegangenen Abschnitten verschiedene Testverfahren zum Vergleich von Sensitivitäten und Spezifitäten und Möglichkeiten für die Adaption der bei-

den Gütemaße auf den Alarmsystem-Kontext vorgestellt wurden, wird im Folgenden eine Vorgehensweise für die Validierung von Alarm-Algorithmen ausgewählt. Der vor Beginn der Datenaufzeichnung festgelegte Plan muss dabei etwas modifiziert und den gegebenen Umständen angepasst werden. Da in der frühen Phase der Studienplanung noch davon ausgegangen wurde, dass auch für das derzeitig eingesetzte Alarmsystem falsch negative Entscheidungen identifiziert werden können, wurde für den Sensitivitätsvergleich der Alarmsysteme A_{alt} und A_{neu} das Nicht-Unterlegenheits-Testproblem

$$H_0^{Se} : Se(A_{alt}) - Se(A_{neu}) \geq \delta \quad \text{vs.} \quad H_1^{Se} : Se(A_{alt}) - Se(A_{neu}) < \delta, \quad \delta = 0.05 \quad (6.23)$$

gemäß (6.4) formuliert, welches mit der Score-Teststatistik $X_S(\delta)$ aus (6.7) behandelt werden sollte. Da für A_{alt} in dieser Studie aber eine Sensitivität von 100% angesetzt werden muss, reduziert sich dieses Testproblem gemäß (6.8) zu

$$H_0^{Se} : Se(A_{neu}) \leq Se_0 \quad \text{vs.} \quad H_1^{Se} : Se(A_{neu}) > Se_0, \quad \text{mit} \quad Se_0 = 0.95, \quad (6.24)$$

für dessen Behandlung auch andere Tests gut geeignet sind.

Für die Überprüfung des Überlegenheits-Testproblems gemäß (6.5) zum Vergleich der Spezifitäten

$$H_0^{Sp} : Sp(A_{alt}) - Sp(A_{neu}) \geq 0 \quad \text{vs.} \quad H_1^{Sp} : Sp(A_{alt}) - Sp(A_{neu}) < 0 \quad (6.25)$$

war ursprünglich der McNemar-Test mit Teststatistik X_M (6.11) vorgesehen. Die Bestimmung der Spezifität sollte dabei auf einer Aufteilung des nicht alarmrelevanten Zeitraums in gleich große Zeitabschnitte basieren. Wie in 6.2.2 diskutiert, ist eine derartige Vorgehensweise für das vorliegende Datenmaterial ungeeignet, da die Fehlalarmhäufigkeit durch den resultierenden Spezifitätswert nicht adäquat reflektiert wird. Stattdessen sollte die in (6.22) vorgeschlagene alternative Spezifitätsdefinition verwendet werden. In diesem Fall ist der klassische McNemar-Test aber nicht anwendbar.

Da möglicherweise Abhängigkeiten zwischen den Alarmsystem-Entscheidungen für einen Patienten vorliegen, ist zudem eine Anpassung der Teststatistiken nötig, da sonst antikonservative Testergebnisse auftreten können. Geeignete Teststatistiken für das Sensitivitäts-Testproblem sind zum einen die mittels des Intraclass-Korrelationskoeffizienten adjustierte Wald-Teststatistik $X_{W'}$ aus (6.13) und zum anderen $X_{W''}$ aus (6.15), bei der die aus der Stichprobentheorie bekannte Varianz des Verhältnisschätzers zur Adjustierung der einfachen Wald-Teststatistik verwendet wird. Verfahren, bei denen die Cluster nicht proportional zu ihrer Größe gewichtet werden und die in Abschnitt 6.1.2 kurz angerissen wurden, werden für diese Studie nicht

berücksichtigt, da alle alarmrelevanten Zustände gleichwertig behandelt werden sollen. Zur Auswahl stehende Teststatistiken für das Spezifitäts-Testproblem sind gegeben durch X_O aus (6.17) und X_R aus (6.18), da diese keine Bestimmung diskordanter und konkordanter Beobachtungspaare erfordern. Marginale Regressionsmodelle werden für den einfachen Spezifitätsvergleich nicht in Betracht gezogen. Diese sollten eher bei weitergehenden Analysen, in denen die Spezifität z.B. abhängig von Kovariablen wie Sedierungsgrad und Schweregrad der Krankheit untersucht wird, zum Einsatz kommen. Da sich die Clustergrößen aufgrund der unterschiedlichen Beobachtungsdauern für die einzelnen Patienten stärker unterscheiden und wenig über die Korrelationsstruktur bekannt ist, wird für die Validierung die Teststatistik X_O verwendet. Bei der Bestimmung von Schätzwerten für die Gesamt-Spezifität wird dabei so vorgegangen, dass zunächst die Spezifitäten für die einzelnen Patienten bestimmt werden. Auf Basis dieser Spezifitäten bzw. der Informationen über die Anzahl richtig negativer Entscheidungen und die Anzahl nicht alarmrelevanter Zustände, wird die Gesamt-Spezifität bestimmt. Dieser Ansatz ist zu wählen, da die Einzel-Spezifitäten für die Teststatistik X_O benötigt werden. Wenn bei einzelnen Patienten der Fall eintritt, dass $FP > FP_{max}$ und somit $\hat{S}_p = 0$, so unterscheidet sich die Gesamt-Spezifität auf Basis der Einzel-Spezifitäten von der Gesamt-Spezifität, die resultiert, wenn alle falsch positiven Alarme und nicht alarmrelevanten Zustände insgesamt betrachtet werden.

Für das Sensitivitäts-Testproblem wird die Teststatistik $X_{W''}$ verwendet, da hier dieselbe Methode zur Berücksichtigung von Abhängigkeiten verwendet wird, wie bei X_O . Gemäß des Intersection-Union Testprinzips wird die Überlegenheit des neuen Alarmsystems in dieser Studie genau dann gezeigt, wenn gilt

$$X_{W''} > u_{1-\alpha} \quad \text{und} \quad X_O < u_\alpha. \quad (6.26)$$

In dieser Studie wird das Signifikanzniveau zu $\alpha = 0.05$ gewählt.

$(1 - \alpha)$ -Konfidenzintervalle für Sensitivitäten werden gemäß $\hat{S}_e \pm u_{\alpha/2} \sqrt{V_C}$ (mit V_C aus (6.14)) bestimmt. Konfidenzintervalle für Spezifitäten erhält man analog.

6.3 Weitere Beurteilungskriterien

Obwohl als Haupt-Beurteilungskriterien Sensitivität und Spezifität festgelegt wurden und ein Vergleich dieser Gütemaße der beiden Alarmsysteme mittels Signifikanztests die Entscheidung hinsichtlich der Eignung eines neuen Alarm-Algorithmus maßgeblich beeinflusst, werden auch andere Eigenschaften der Alarmsysteme betrachtet. Diese

werden im Folgenden vorgestellt.

An erster Stelle ist die absolute Anzahl generierter Alarme zu nennen. Ein Vergleich der Alarmhäufigkeiten von altem und neuen Alarm-Algorithmus kann einen ersten Hinweis auf die Eignung des neuen Verfahrens liefern. Ein neues System, welches mehr Alarme als das derzeitige verwendete System auslöst, ist bei gleicher Sensitivität in jedem Fall ungeeignet, da eine Motivation für die Entwicklung neuer Alarm-Algorithmen gerade die Reduzierung überflüssiger Alarme ist. Diese Anzahlen sollten nicht nur insgesamt, sondern auch für die einzelnen Patienten verglichen werden. Selbst bei einer niedrigeren Gesamtanzahl Alarme, kann bei einzelnen Patienten die Anzahl Alarme des neuen Alarm-Algorithmus größer sein. Eine Untersuchung dieser Fälle gibt möglicherweise Aufschluss über Ursachen hierfür. Neben der Anzahl Alarme insgesamt, kann auch die Anzahl Alarme pro Zeiteinheit, z.B. pro Stunde, betrachtet werden.

Auch die Anzahlen falsch positiver Alarme insgesamt bzw. pro Zeiteinheit sind sinnvolle Beurteilungskriterien, insbesondere vor dem Hintergrund der begrenzten Aussagekraft von Schätzwerten für die Spezifität (siehe Abschnitt 6.2.2).

Ein im Gebiet der Evaluierung diagnostischer Verfahren weit verbreitetes Gütemaß ist der sogenannte Positive Prädiktive Wert (auch: Vorhersagewert, PPV). Dieser gibt im Falle eines Alarms die Wahrscheinlichkeit dafür an, dass ein alarmrelevanter Zustand vorliegt und ist definiert als

$$\text{PPV} = P(\text{Zustand alarmrelevant} | \text{Alarm}). \quad (6.27)$$

Diese Kenngröße ist insbesondere für die praktische Anwendung eines Alarmsystems von Bedeutung. Bei Kenntnis des PPV kann der Mediziner im konkreten Falle eines Alarms einschätzen, wie sicher dieser auf einen alarmrelevanten Zustand hinweist und kann entsprechend reagieren. Ein Nachteil des PPV ist seine Prävalenz-Abhängigkeit. Ist die Wahrscheinlichkeit für das Auftreten eines alarmrelevanten Zustandes gering, ist auch der Vorhersagewert klein.

Weiteren Aufschluss über die Funktionsweise des neuen Alarm-Algorithmus kann eine Analyse hinsichtlich der Alarmsituationen liefern, in denen hinweisende Alarme auftreten. Für die Sensitivitäts- und Spezifitätsbestimmung werden die Annotationen nur ausgenutzt, um alarmrelevante Situationen zu identifizieren. Die Sensitivität bezieht sich auch nur auf die Erkennung dieser Situationen. Alarme außerhalb solcher Situationen werden für die Spezifitätsbestimmung als falsch positiv angesehen. Viele dieser Alarme tragen allerdings die Annotation „hinweisend“ und sind somit zwar nicht dringend notwendig aber informativ. Für das neue Alarmsystem sollte somit untersucht werden, ob in Situationen mit hinweisenden Alarmen des einfachen Schwellwertalarm-systems auch neue Alarme auftreten. Insbesondere könnte möglicherweise gezeigt wer-

den, dass der neue Algorithmus diese Situationen mit weniger Alarmen anzeigt.

In Abschnitt 2.3 wird von zwei Studien berichtet, in denen Alarm-Algorithmen untersucht werden, die ein ähnliches Ziel wie Signalextraktions-Alarme verfolgen. In diesen Studien erfolgt ebenfalls ein Vergleich zwischen dem derzeitigen und dem jeweils neu entwickelten Verfahren anhand klinischer Daten. Der Vergleich ist allerdings in beiden Fällen nicht konfirmatorisch, sondern rein deskriptiv: Mäkivirta et al. (1991) verwenden zur Bewertung der Alarmsysteme die Sensitivität, den Anteil korrekter Alarme an allen Alarmen (PPV) sowie die Häufigkeit falsch positiver Alarme pro Stunde.

Charbonnier und Gentil (2007) wählen folgende Vorgehensweise zur Beurteilung ihres neuen Online-Segmentierung-Verfahrens: Zunächst wird für jeden Alarm des alten Alarmsystems bestimmt, ob dieser gleichzeitig mit einem Alarm des neuen Systems auftritt. Hierfür wird für die Alarme beider Systeme auf beiden Seiten der Alarmperiode (Beginn bis Ende des Alarms) ein Zeitfenster der Länge 30 Sekunden hinzugefügt. Überlappt sich das so entstandene Zeitintervall für einen alten Alarm mit dem Intervall eines neuen Alarms, so wird gesagt: er tritt gleichzeitig mit dem neuen Alarm auf (Notation: $A_{alt}-A_{neu}$). Gibt es für einen alten Alarm keine Intervall-Überlappung mit einem neuen Alarm, so gilt der alte Alarm als vom neuen System *unterdrückt* ($A_{alt}-\overline{A_{neu}}$). Dieses Schema ist in Abbildung 6.6(a) dargestellt. Dieselbe Prozedur wird dann ausgehend von den neuen Alarmen durchgeführt, allerdings mit einem kleineren Zeitfenster um die Alarmperiode (15 Sekunden). Neue Alarme, bei denen keine Intervall-Überlappung mit einem alten Alarm auftritt, werden als *zusätzliche* Alarme angesehen ($A_{neu}-\overline{A_{alt}}$). Die beiden Anzahlen gleichzeitiger Alarme ($A_{alt}-A_{neu}$ und $A_{neu}-A_{alt}$) können unterschiedlich groß sein, da z.B. während eines länger andauernden neuen Alarms mehrere kurz andauernde alte Alarme auftreten können. Die Zeitfenster um die Alarmperioden werden unterschiedlich groß gewählt, um die Anzahl unterdrückter Alarme zu minimieren

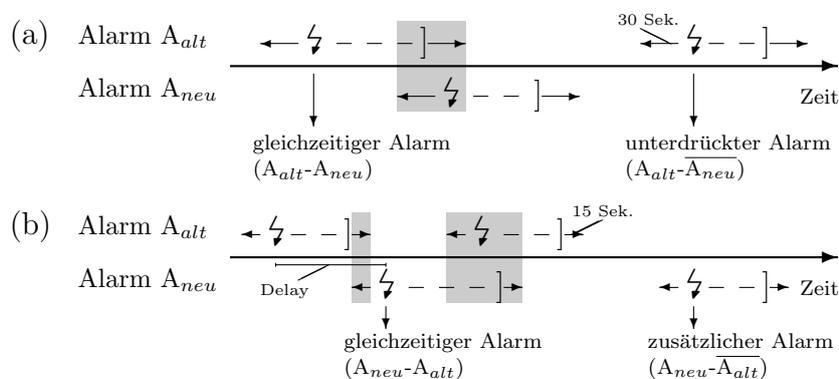


Abbildung 6.6: Bestimmung unterdrückter alter und zusätzlicher neuer Alarme (Zeichnung in Anlehnung an Abbildung 7 aus Charbonnier und Gentil (2007))

und die Anzahl zusätzlicher Alarme zu maximieren. Zusätzlich wird die Verzögerung (Delay) zwischen zwei gleichzeitigen Alarmen bestimmt (siehe Abbildung 6.6(b)). Auf Basis von Annotationen kann für die unterdrückten Alarme überprüft werden, ob diese klinisch relevante Situationen anzeigen oder ob eine Unterdrückung dieser Alarme keinen Sensitivitätsverlust bedeutet. Für die zusätzlich generierten Alarme müssen die Ursachen gefunden werden.

Kapitel 7

Der Referenzdatensatz

Die Datenaufzeichnung für die Klinische Studie erfolgte zwischen Januar 2006 und Mai 2007. Insgesamt wurden Daten für 68 verschiedene „Fälle“ erhoben. Dabei kann es sich bei verschiedenen Fällen durchaus um denselben Patienten handeln, dann aber in unterschiedlichen Krankheitsstadien.

7.1 Beobachtungsdauer und Alarmhäufigkeit

Insgesamt umfasst der Referenzdatensatz einen Beobachtungszeitraum von 982 Stunden. Die Beobachtungsdauern für die einzelnen in den Referenzdatensatz aufgenommenen Fälle sind in Tabelle 7.1 angegeben. Die durchschnittliche Beobachtungsdauer beträgt 14.4 Stunden (Standardabweichung 7.5 Std.). Für den Fall mit der kürzesten Beobachtungsdauer liegen Daten für 55 Minuten vor. Die längste Aufzeichnung deckt einen Zeitraum von über 25 Stunden ab. Die Beobachtungszeiträume sind nicht immer zusammenhängend, sondern können für einige Minuten oder Stunden unterbrochen sein. Dies tritt z.B. auf, wenn der Patient für bestimmte Untersuchungen für einige Zeit vom Monitor getrennt werden muss.

Tabelle 7.1 enthält außerdem die aufsummierte Dauer der Zeiträume, in denen die Alarmfunktion mittels der „Alle Alarme Aus“-Taste (siehe Abschnitt 4.1.1) deaktiviert wurde. Insgesamt machen die Alarm-Deaktivierungszeiträume ungefähr 8% der Beobachtungszeit aus.

Fall	Dauer	Alarmer deaktiviert	Alarmer aktiviert	Anz. Alarmer	Alarmer/ Std.	Anz. Schwell- wertalarmer	Schwellwert- alarmer/Std.
1	03:00:01	00:27:00	02:33:01	32	11	24	8
2	10:44:53	01:14:54	09:29:59	55	5	43	4
3	14:28:14	00:22:08	14:06:06	49	3	7	0
4	20:14:44	00:16:33	19:58:11	107	5	19	1
5	00:55:01	00:11:15	00:43:46	7	8	2	2
6	09:13:07	01:09:36	08:03:31	121	13	65	7
7	24:10:43	01:00:15	23:10:28	28	1	11	0
8	19:56:10	01:08:09	18:48:01	43	2	29	1
9	14:33:37	02:01:56	12:31:41	418	29	361	25
10	14:20:57	01:18:44	13:02:13	289	20	209	15
11	08:04:22	00:18:00	07:46:22	234	29	184	23
12	09:21:30	00:37:30	08:44:00	48	5	13	1
13	02:52:05	00:02:00	02:50:05	7	2	2	1
14	13:47:54	01:09:01	12:38:53	155	11	89	6
15	08:54:52	00:36:00	08:18:52	38	4	32	4
16	20:27:28	00:51:29	19:35:59	298	15	283	14
17	05:13:00	00:12:00	05:01:00	40	8	21	4
18	17:51:56	01:11:44	16:40:12	59	3	35	2
19	06:53:12	01:08:42	05:44:30	69	10	37	5
20	04:09:33	00:03:01	04:06:32	31	7	31	7
21	05:29:30	00:20:27	05:09:03	41	7	39	7
22	05:55:28	00:21:00	05:34:28	37	6	15	3
23	11:13:46	00:54:29	10:19:17	46	4	27	2
24	22:18:10	00:46:03	21:32:07	56	3	22	1
25	24:00:00	02:17:01	21:42:59	60	2	27	1
26	16:31:44	01:14:33	15:17:11	76	5	22	1
27	24:42:01	02:15:26	22:26:35	104	4	94	4
28	23:56:34	02:04:58	21:51:36	126	5	103	4
29	24:56:54	03:10:01	21:46:53	165	7	107	4
30	11:44:41	03:07:19	08:37:22	118	10	80	7
31	22:42:54	02:00:51	20:42:03	183	8	115	5
32	22:51:51	03:24:33	19:27:18	197	9	160	7
33	06:18:20	00:11:29	06:06:51	29	5	6	1
34	08:07:21	01:47:38	06:19:43	127	16	88	11
35	05:09:55	00:13:36	04:56:19	8	2	4	1
36	15:41:55	02:12:58	13:28:57	92	6	55	4
37	14:58:23	01:25:49	13:32:34	61	4	41	3
38	16:21:59	00:58:20	15:23:39	44	3	19	1
39	17:18:07	00:27:00	16:51:07	28	2	10	1
40	22:13:45	01:11:34	21:02:11	97	4	43	2
41	17:24:16	00:38:36	16:45:40	16	1	5	0
42	19:23:53	00:32:10	18:51:43	21	1	14	1
43	24:11:03	00:43:58	23:27:05	35	1	15	1
44	22:57:58	02:23:22	20:34:36	116	5	76	3
45	10:49:26	01:04:25	09:45:01	37	3	20	2
46	25:15:48	03:11:56	22:03:52	118	5	89	4
47	25:17:15	01:24:44	23:52:31	28	1	19	1
48	17:34:58	01:26:39	16:08:19	79	4	45	3
49	01:16:40	00:05:15	01:11:25	1	1	1	1
50	09:09:50	00:52:58	08:16:52	23	3	18	2
51	11:40:28	00:38:22	11:02:06	43	4	39	3
52	24:00:00	01:15:39	22:44:21	24	1	5	0
53	07:15:53	00:19:22	06:56:31	51	7	32	4
54	01:19:01	00:03:00	01:16:01	8	6	1	1
55	18:38:05	01:08:26	17:29:39	167	9	129	7
56	07:10:41	00:35:36	06:35:05	27	4	20	3
57	23:38:21	02:13:06	21:25:15	145	6	102	4
58	15:35:45	00:22:27	15:13:18	39	3	31	2
59	06:00:48	00:29:45	05:31:03	14	2	11	2
60	08:08:50	00:33:01	07:35:49	29	4	25	3
61	01:58:42	00:06:00	01:52:42	21	11	9	5
62	19:45:00	01:45:44	17:59:16	160	8	113	6
63	21:43:31	02:05:53	19:37:38	127	6	103	5
64	23:24:15	02:10:10	21:14:05	481	21	417	18
65	14:16:41	01:01:59	13:14:42	137	10	106	7
66	09:07:59	00:29:52	08:38:07	60	7	51	6
67	19:06:04	00:51:46	18:14:18	15	1	13	1
68	24:14:32	01:39:25	22:35:07	89	4	63	3
Insg.	982:12:20	75:58:38	906:13:42	5934	6	4146	4

Tabelle 7.1: Beobachtungsdauer, Alarm-Deaktivierungszeit, Alarm-Aktivierungszeit, Anzahl Alarmer, Anzahl Alarmer pro Stunde (gerundet), Anzahl Schwellwertalarmer und Anzahl Schwellwertalarmer pro Stunde (gerundet) für den Referenzdatensatz

In den aufgezeichneten 982 Stunden wurden insgesamt 5934 Alarme ausgelöst. Dies entspricht einer Frequenz von 6 Alarmen pro Stunde. Bei 70% der aufgetretenen Alarme handelt es sich um Schwellwertalarme. Für die einzelnen Fälle ist die Anzahl der Alarme bzw. Schwellwertalarme pro Stunde allerdings teilweise sehr unterschiedlich. Für die vier Tageszeiten Morgen (6-12 Uhr), Nachmittag (12-18 Uhr), Abend (18-0 Uhr) und Nacht (0-6 Uhr) lässt sich feststellen, dass diese bezüglich der Beobachtungsdauer ungefähr gleich stark vertreten sind. Der größte Anteil entfällt mit 27% auf den Nachmittag, der kleinste Anteil mit vier Prozentpunkten weniger auf die Nachtstunden (siehe Tabelle B.1 im Anhang). Allerdings wurden bei nur jeweils neun Fällen keine Daten für morgens oder nachmittags erhoben, während für den Abend bei 18 und für die Nacht bei 26 Fällen keine Daten vorliegen. Für 33 der 68 beobachteten Fälle gibt es Aufzeichnungen für alle vier Tageszeiten, bei vier Fällen beschränken sich die Beobachtungen auf eine einzige Tageszeit.

Bezüglich der Anzahl der Alarme und der Anzahl Schwellwertalarme pro Stunde gibt es Unterschiede zwischen den Tageszeiten. Morgens und nachmittags ist die Alarmfrequenz höher als abends und nachts, was unter anderem dadurch zu erklären ist, dass tagsüber häufiger Pflegemaßnahmen am Patienten vorgenommen werden.

7.2 Vitalparameter

Bevor eine ausführliche deskriptive Analyse der Alarmmeldungen und Annotationen stattfindet, wird zunächst festgehalten, welche Vitalparameter für die im Referenzdatensatz vorliegenden Fälle während der Datenaufzeichnung gemessen wurden und für welche Vitalparameter die Alarmfunktion aktiviert war.

Tabelle B.2 im Anhang zeigt, dass die Vitalparameter Herzfrequenz, Puls, Sauerstoffsättigung, systolischer, mittlerer und diastolischer arterieller Blutdruck bei allen 68 Fällen sekundlich bestimmt wurden. Auch die Arrhythmieerkennung war in allen Fällen eingestellt. Weitere häufig gemessene Variablen sind PVC/min (die Anzahl vorzeitiger ventrikulärer Kontraktionen pro Minute), die Bluttemperatur (Ta), die Atemfrequenz sowie verschiedene ST-Segmente. Alle weiteren in Tabelle B.2 aufgeführten Vitalparameter wurden nur vereinzelt gemessen.

Eine Auflistung der Vitalparameter, für die bei den einzelnen Fällen die Alarmfunktion aktiviert war, findet sich in Tabelle B.3 im Anhang. Arrhythmien, mittlerer arterieller Blutdruck, Herzfrequenz und die Sauerstoffsättigung wurden für alle Fälle im Referenzdatensatz per Alarmsystem überwacht. Der systolische arterielle Blutdruck wurde

in nur einem Fall nicht überwacht. Für PVC/min war die Alarmfunktion für 63 der 68 Fälle aktiviert. Nur für acht Fälle wurde die Atemfrequenz überwacht, obwohl dieser Vitalparameter bei 37 Fällen gemessen wurde. Da der Blutdruck für alle Fälle im Referenzdatensatz invasiv gemessen wurde, fand eine Messung und Überwachung des nicht-invasiven systolischen und mittleren Blutdrucks entsprechend nur in wenigen Ausnahmefällen statt. Eine Überwachung des mittleren pulmonalarteriellen Druckes erfolgte bei vier Fällen. Auffällig ist, dass Puls und diastolischer arterieller Blutdruck zwar bei allen Patienten gemessen, aber in keinem Fall per Alarmsystem überwacht wurden.

7.3 Alarmmeldungen

Im Folgenden werden die vom Infinity Monitoring System[®] generierten Alarme näher hinsichtlich Alarmgrad, Alarmart, auslösendem Vitalparameter und Ursache analysiert. Jeder ausgelöste Alarm ist mit einem Alarmgrad versehen. Ein Alarm trägt entweder den Grad „hinweisend“ (ADV), „ernst“ (SER) oder „lebensbedrohlich“ (LT) (siehe auch Abschnitt 4.1.1). Im Referenzdatensatz liegen insgesamt 5102 (86%) ernste Alarme, 826 (14%) hinweisende Alarme und 6 (0.1%) lebensbedrohliche Alarme vor.

Anhand der Alarmmeldungen können grob vier Arten von Alarmen unterschieden werden (siehe Tabelle 7.2). Schwellwertalarme treten dabei am häufigsten auf. Ihr Anteil an allen ausgelösten Alarmen beträgt knapp 70%. Dabei ist in 2341 Fällen eine Unterschreitung der unteren Alarmgrenze und in 1805 Fällen eine Überschreitung der oberen Alarmgrenze die Alarmursache. Bei allen aufgetretenen Schwellwertalarmen handelt es sich um Alarme mit dem Alarmgrad „ernst“.

Eine weitere Gruppe von Alarmmeldungen wird im Folgenden unter dem Sammelbegriff „sonstige ernste Alarme“ zusammengefasst. Hierbei handelt es sich um Meldungen mit Alarmgrad SER, die einen statischen invasiven Blutdruck, Blutdruckwerte außerhalb

Alarmart	Anzahl	Anteil
Schwellwertalarme	4146	0.6987
Sonstige ernste Alarme	873	0.1471
Technische Alarme	807	0.1360
Arrhythmie-Alarme	108	0.0182
Insgesamt	5934	1.0000

Tabelle 7.2: Alarmarten

des Messbereichs oder eine Koinzidenz der Atemfrequenz mit der Herzfrequenz anzeigen. Ungefähr 15% aller Alarme tragen diese Alarmmeldungen. Eine Übersicht über die Häufigkeiten der einzelnen Meldungen findet sich in Tabelle B.4. Es ist zu bemerken, dass nicht alle ernstesten Alarme, die keine Schwellwertalarme sind, in die Gruppe der sonstigen ernstesten Alarme fallen. Einige Arrhythmie-Alarme tragen ebenfalls den Alarmgrad „ernst“, werden aber gesondert betrachtet.

Unter dem Begriff „Technische Alarme“ sind alle Alarme mit Alarmgrad ADV zusammengefasst, bei denen es sich nicht um Arrhythmie-Alarme handelt. Technische Alarme werden hauptsächlich durch Signalprobleme bei der Messung der Sauerstoffsättigung verursacht (siehe Tabelle B.5) und machen fast 14% aller Alarme aus.

Die vierte Alarmgruppe beinhaltet Alarme zu Arrhythmie-Ereignissen. Knapp 2% der Alarme des Referenzdatensatzes sind Arrhythmie-Alarme. Welche Arrhythmien alarmiert werden, ist in Tabelle B.6 zusammengefasst. In der Gruppe der Arrhythmie-Alarme treten alle drei möglichen Alarmgrade auf (SER: 83, ADV: 19, LT: 6).

Insgesamt werden die meisten Alarme durch den systolischen arteriellen Blutdruck ausgelöst (siehe Tabelle 7.3). Es folgen Alarme für die Sauerstoffsättigung, die Herzfrequenz und den mittleren arteriellen Blutdruck. Für die in dieser Studie besonders interessierenden Schwellwertalarme zeigt sich eine ähnliche Rangfolge der alarmanzulösenden

Vitalparameter	Anz. Alarme	Anteil	Anz. Schwellwertalarme	Anteil an Schwellwertalarmen	Anteil an allen Alarmen
ART.S	2605	0.4390	1857	0.4479	0.3129
SpO2	1517	0.2556	784	0.1891	0.1321
HR	789	0.1330	727	0.1753	0.1225
ART.M	503	0.0848	485	0.1170	0.0817
RESP	196	0.0330	163	0.0393	0.0275
ARR	103	0.0174	-	-	-
PA.M	89	0.0150	53	0.0128	0.0089
ART.S (2)	51	0.0086	33	0.0080	0.0056
ART.M (2)	51	0.0086	24	0.0058	0.0040
NBP.S	14	0.0024	7	0.0017	0.0012
PVC/min	12	0.0020	12	0.0029	0.0020
PA.D	3	0.0005	0	0.0000	0.0000
NBP.M	1	0.0002	1	0.0002	0.0002
Insgesamt	5934	1.0000	4146	1.0000	0.6987

Tabelle 7.3: Alarmanzulösende Vitalparameter für alle Alarme insgesamt und für Schwellwertalarme. *ART.S (2) und ART.M (2) bedeutet, dass parallel zu einer bereits laufenden arteriellen Drucküberwachung zusätzlich eine zweite durchgeführt wurde.*

Vitalparameter (siehe Tabelle 7.3). Durch den systolischen arteriellen Blutdruck werden die meisten Schwellwertalarme ausgelöst, wieder gefolgt von Sauerstoffsättigung, Herzfrequenz, mittlerem arteriellen Druck und Atemfrequenz.

Eine Rangfolge aller ausgelösten unterschiedlichen Alarmmeldungen – d.h. Vitalparameter plus Ursache des Alarms – findet sich in Tabelle B.7 im Anhang. Die am häufigsten auftretenden Alarme sind untere und obere Schwellwertalarme für den systolischen arteriellen Blutdruck. Weitere sehr häufig auftretende Alarmmeldungen sind untere Schwellwertverletzungen für die Sauerstoffsättigung und obere Schwellwertverletzungen für die Herzfrequenz.

7.4 Annotationen

Die vom Monitoring System während der Datenaufzeichnung generierten Alarme wurden anhand verschiedener Kriterien bewertet. Es erfolgte eine Beurteilung der technischen Validität, eine Klärung der Frage, ob der Alarm durch eine Manipulation ausgelöst wurde und eine Entscheidung über die Alarmrelevanz der Alarmsituation (siehe auch Abschnitt 5.2). Diese Annotationen werden im Folgenden analysiert.

Zunächst werden die Annotationen für alle ausgelösten Alarme insgesamt betrachtet. Abbildung 7.1 veranschaulicht die Häufigkeiten der verschiedenen Bewertungen. Die zugehörigen Anzahlen und Anteile finden sich in Tabelle B.8. Es zeigt sich, dass ein

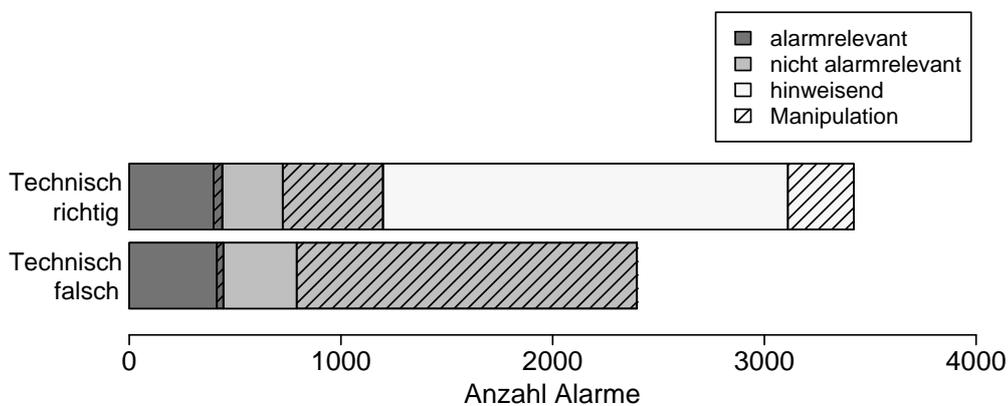


Abbildung 7.1: Häufigkeiten der Alarm-Annotationen für alle ausgelösten Alarme

relativ hoher Anteil durch technisch falsche Alarme zustande kommt, die entweder auf Basis einer inkorrekten Messung gegeben werden oder die das vorliegende Problem durch eine falsche Meldung anzeigen. Die häufigsten Ursachen für technisch falsche Alarme sind Signalprobleme, Lagerung des Patienten durch das medizinische Personal, Bewegungen des Patienten und Blutabnahme (siehe Tabellen B.9 und B.10). 58% aller Alarme sind technisch richtig. Für 55 Alarme wurde in den Annotationen festgehalten, dass sie nicht beurteilt werden konnten. 59 Alarme wurden – bedingt durch das zeitweilig gehäufte Auftreten von Alarmen kurz hintereinander – beim Annotieren übersehen und nicht bewertet, so dass für 2% aller Alarme keine Einschätzung der Alarmsituation vorliegt. In Abbildung 7.1 sind diese Alarme ohne Bewertung nicht berücksichtigt.

Der Anteil der Alarme, die während eines alarmrelevanten Zustandes ausgelöst wurden, liegt bei 15%. Dabei handelt es sich bei der einen Hälfte dieser Alarme um alarmrelevante Situationen, die sich auf den Gesundheitszustand des Patienten beziehen, d.h. die Alarme haben die Annotation technisch richtig. Bei der anderen Hälfte handelt es sich um alarmrelevante Situationen aufgrund eines technischen Problems.

46% aller Alarme werden in Situationen gegeben, die keinen Alarm erfordern. Dabei wurden 77% dieser Alarme durch eine Manipulation verursacht – d.h. diese Fehlalarme sind nur bedingt dem Monitorsystem anzulasten. Ein ebenfalls großer Anteil von 37% aller ausgelösten Alarme wurde in den Annotationen als hinweisend bewertet, d.h. diese Alarme bieten Informationsgehalt für den Mediziner, erfordern aber keine unmittelbare diagnostische oder therapeutische Reaktion.

Der Anteil der durch eine Manipulation verursachten Alarme beträgt insgesamt 42%. Nur für 3% dieser Alarme wurde die Situation als alarmrelevant beurteilt. Der Anteil manipulationsbedingter Alarme mit Annotation „nicht alarmrelevant“ beträgt dagegen beträchtliche 84%. Zum Vergleich: Für alle Alarme ohne Manipulation beträgt letztgenannter Anteil nur 19%.

Durch Manipulationen ausgelöste Alarme werden in 84% der Fälle durch eine Manipulation am Monitorsystem verursacht. Manipulationen am Patienten – zu denen z.B. auch die Gabe von Medikamenten gezählt wird – sind nur in 16% der Fälle für den Alarm verantwortlich. Abbildung B.1 zeigt, dass die durch Manipulationen am Monitorsystem ausgelösten Alarme meist technisch falsch sind. Dagegen sind durch Manipulationen am Patienten ausgelöste Alarme in 98% der Fälle technisch richtig. Dies ist dadurch zu erklären, dass durch Manipulationen am Monitorsystem – etwa bei einer Blutabnahme – häufig keine korrekte Messung mehr möglich ist. Bei Manipulationen am Patienten handelt es sich aber hauptsächlich um Medikamentengaben, die eine tatsächliche Änderung der Messwerte hervorrufen. Für Alarme aufgrund von

Manipulationen am System wurde die Situation in den meisten Fällen als nicht alarmrelevant bewertet. Alarme aufgrund von Manipulationen am Patienten sind meist als hinweisend annotiert.

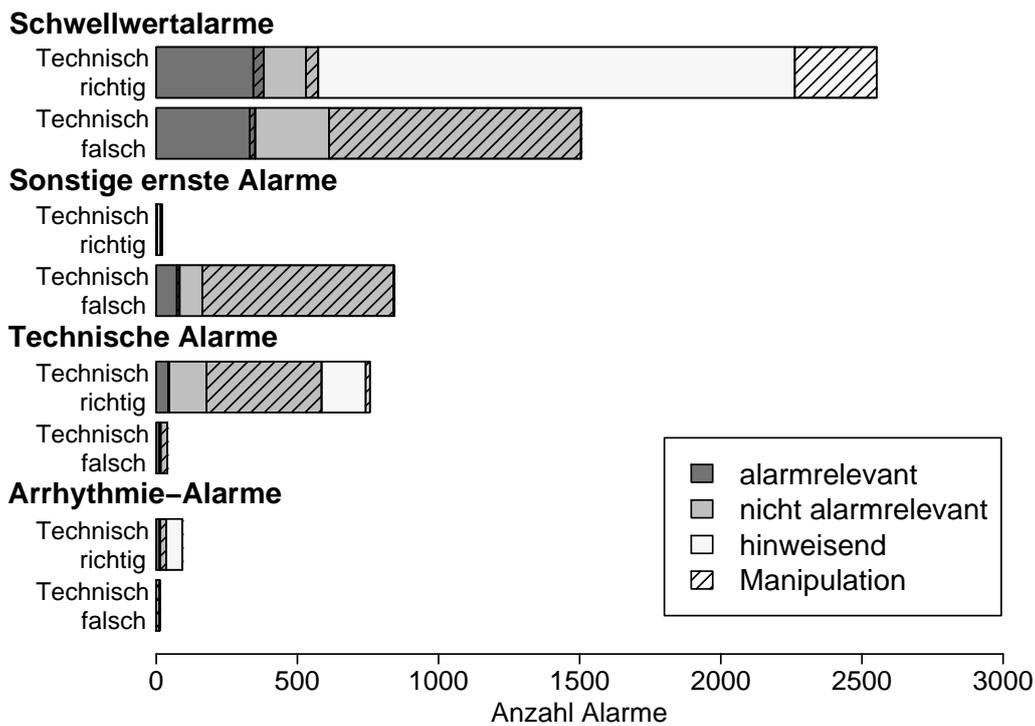


Abbildung 7.2: Häufigkeiten der Alarm-Annotationen für Schwellwertalarme, sonstige ernste (SER) Alarme (d.h. ohne Schwellwertalarme und Arrhythmiealarme), technische Alarme und Arrhythmie-Alarme

Die Häufigkeitsverteilungen der verschiedenen Kombinationen von technischer Validität, Manipulation und Alarmrelevanz unterscheiden sich für die vier Alarmarten (siehe Abbildung 7.2 und Tabellen B.11-B.14).

Bezüglich der technischen Validität zeigt sich, dass 62% aller Schwellwertalarme auf korrekten Messwerten basieren; 36% der Schwellwertalarme sind dagegen technisch falsch und können z.B. durch Ausreißer bedingt sein. In der Gruppe der sonstigen ernstesten Alarme wurden fast alle Alarme als technisch falsch beurteilt. Bei den technischen Alarmen weist die Alarmmeldung in 94% der Fälle korrekt auf das vorliegende technische Problem hin. Arrhythmie-Alarme sind ebenfalls mehrheitlich (84%) technisch richtig.

Unter den vier Alarmarten ist für Schwellwertalarmlarmer der Anteil der als alarmrelevant bewerteten Zustände mit 18% am größten. Allerdings sind nur bei 9% aller Schwellwertalarmlarmer sowohl Alarmrelevanz als auch technische Richtigkeit gegeben. Bei den übrigen Fällen bezieht sich die Alarmrelevanz nicht auf den Gesundheitszustand des Patienten, sondern auf ein technisches Problem. In diesen Situationen ist zwar ein Alarm erforderlich, aber die Meldung, dass eine Alarmgrenze überschritten wurde, ist inadäquat und für das medizinische Personal irreführend. Für 32% der Schwellwertalarmlarmer ist der Zustand zum Zeitpunkt des Alarms nicht alarmrelevant. 70% dieser Alarmlarmer wurden durch eine Manipulation herbeigeführt. Fast die Hälfte aller Schwellwertalarmlarmer wurden als hinweisend annotiert.

Bei den sonstigen ernstesten Alarmlarmer sind 10% mit alarmrelevant annotiert, wobei es sich in allen Fällen um technisch falsche Alarmlarmer handelt. Für einen sehr großen Anteil von 89% ist die Alarmlarmerituation nicht alarmrelevant. Die Gruppe der technischen Alarmlarmer weist mit 7% den niedrigsten Anteil Alarmlarmer mit alarmrelevantem Zustand auf. Arrhythmie-Alarmlarmer sind in der Mehrzahl als hinweisend annotiert.

Da in dieser Arbeit ein besonderes Augenmerk auf Schwellwertalarmlarmer gelegt wird, werden die Annotationen auch getrennt für die in Bezug auf Schwellwertalarmlarmer wichtigsten Vitalparameter untersucht. Über 95% aller Schwellwertalarmlarmer sind auf die fünf Vitalparameter systolischer arterieller Blutdruck, Sauerstoffsättigung, Herzfrequenz, mittlerer arterieller Blutdruck und Atemfrequenz zurückzuführen. Die Häufigkeiten der unterschiedlichen Annotationen für diese Parameter sind in Abbildung 7.3 grafisch dargestellt, wobei Alarmlarmer für die beiden arteriellen Blutdrücke zusammengefasst sind. In den Tabellen B.15-B.18 finden sich die genauen Anzahlen und Anteile. Auch hier zeigen sich deutliche Unterschiede der Häufigkeitsverteilungen bezüglich technischer Validität, Manipulation und Alarmrelevanz. Für arterielle Drücke und Herzfrequenz ist die Mehrzahl der Alarmlarmer technisch richtig, für die Sauerstoffsättigung und die Atemfrequenz ist der größte Teil der Alarmlarmer technisch falsch. Der Anteil der Alarmlarmer, die einen alarmrelevanten Zustand anzeigen, ist für die Sauerstoffsättigung mit 35% sehr groß, allerdings sind über 60% dieser Alarmlarmer technisch falsch. D.h. in diesen insgesamt 172 Fällen hätte die Alarmlarmermeldung nicht auf eine Schwellwertverletzung sondern auf ein technisches Problem hinweisen müssen. Für die arteriellen Blutdrücke beträgt der Anteil der Alarmlarmer in alarmrelevanten Situationen 17%. Aber auch hier ist in 40% der Fälle ein Schwellwertalarm unangebracht. Mit einem Anteil von 29% sind relativ wenige der arteriellen Blutdruckalarmlarmer als nicht alarmrelevant annotiert. Von diesen Alarmlarmer wurden 80% durch eine Manipulation herbeigeführt. 52% aller Schwellwertalarmlarmer für systolischen und mittleren arteriellen Blutdruck sind als hinweisend bewertet. Nur für

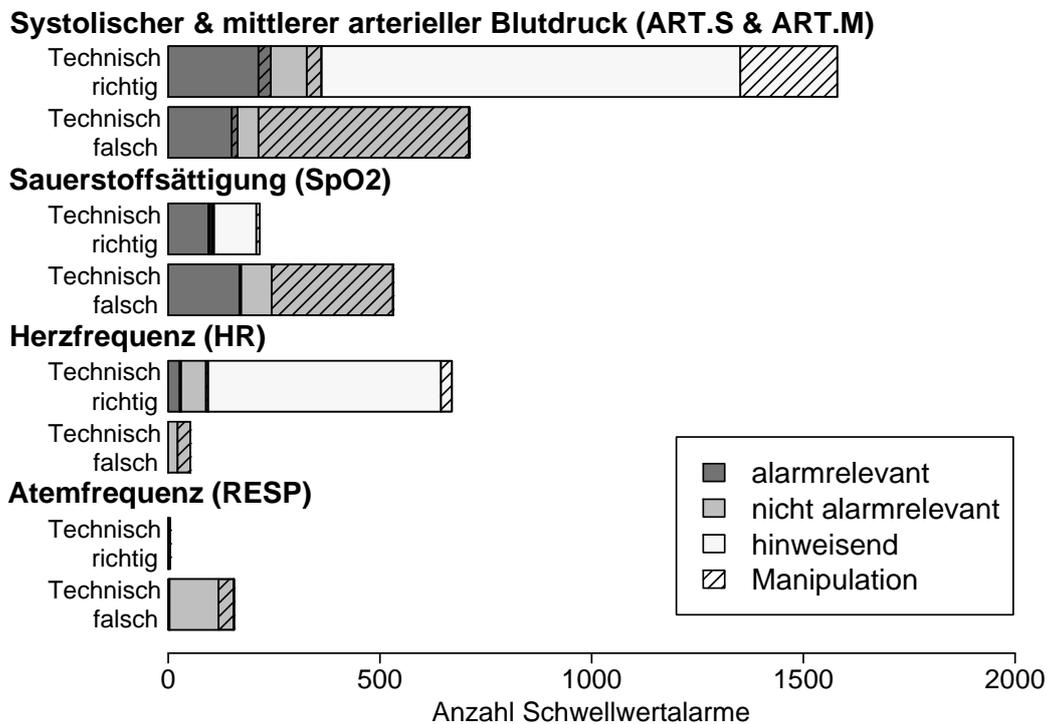


Abbildung 7.3: Häufigkeiten der Alarm-Annotationen für die Schwellwertalarmlarmer der wichtigsten Vitalparameter

4% aller Herzfrequenzalarmlarmer wurde die Situation zum Alarmzeitpunkt als alarmrelevant beurteilt. Der überwiegende Anteil HR-Alarmlarmer (79%) ist als hinweisend annotiert. Alarmlarmer für die Atemfrequenz treten fast ausschließlich (93%) in nicht alarmrelevanten Situationen auf.

Durch Manipulationen ausgelöste Alarmlarmer kommen bei den arteriellen Blutdrücken und der Sauerstoffsättigung häufiger vor als bei Atemfrequenz und Herzfrequenz. Bei letzterer beträgt der Anteil nur 9%, während bei der Sauerstoffsättigung der entsprechende Anteil 39% beträgt.

Die Analyse der Annotationen für die Schwellwertalarmlarmer zeigt, dass viele der außerhalb von alarmrelevanten Situationen aufgetretenen Alarmlarmer entweder durch Manipulationen verursacht werden oder, im Falle der sehr stark vertretenen hinweisenden Alarmlarmer, durch Grenzverletzungen, die zwar erhöhte Aufmerksamkeit aber keine unmittelbare Handlung erfordern. Nicht manipulationsbedingte Ausreißer sind eher selten Ursache unerwünschter Alarmlarmer, nicht zuletzt wegen der für die meisten Schwellwertalarmlarmer gel-

Parameter	Grenze	Alarmer in Gruppen		
		Anzahl	Anteil	Insgesamt
ART.S/ART.M	Obere	701	0.67	1051
	Untere	740	0.57	1291
SpO2	Untere	331	0.42	784
HR	Obere	389	0.69	560
	Untere	96	0.57	167
RESP	Obere	65	0.42	153
	Untere	2	0.20	10

Tabelle 7.4: Anzahl und Anteil von Alarmen für die der Abstand zwischen dem Anfang des Alarms und dem Ende des letzten Alarms gleicher Alarmursache oder der Abstand zwischen dem Ende des Alarms und dem Anfang des nächsten Alarms gleicher Alarmursache nicht mehr als 60 Sekunden beträgt.

tenden Alarmbestätigungsfrist von einigen Sekunden (siehe Tabelle 4.1 in Abschnitt 4.1.1). Unter anderem die Vermeidung solcher Alarme liefert aber eine wichtige Motivation für neue, auf robuster Signalextraktion basierende Alarme. Dennoch bietet die Signalextraktion Potential für die Verbesserung des derzeitigen Schwellwertsystems. Diesbezüglich lohnt sich ein genauere Blick auf das Auftretensmuster von Alarmen – insbesondere auf die Häufung von Alarmen gleicher Alarmursache kurz hintereinander. Tabelle 7.4 zeigt für die einzelnen Alarmursachen die Anzahl und den Anteil der Alarme, für die der nächste Alarm gleicher Meldung in einem Abstand von nicht mehr als 60 Sekunden vorkommt. Für die arteriellen Blutdrücke und die Herzfrequenz treten über die Hälfte der Schwellwertalarme in Gruppen auf, bei der Sauerstoffsättigung und der Atemfrequenz ist der Anteil etwas geringer. Ursachen für solche Alarmgruppen können z.B. kurzfristige Schwankungen der Originalbeobachtungen um die Alarmgrenze sein. In solchen Fällen reicht häufig ein Alarm aus, um auf die Situation aufmerksam zu machen. Mit Hilfe eines auf robuster Signalextraktion basierenden Alarm-Algorithmus kann die Anzahl redundanter Alarme für eine Alarmsituation möglicherweise verringert werden, da das Signal die kurzfristigen Schwankungen um die Alarmgrenze nicht nachzeichnet.

Kapitel 8

Regeln für die Offline-Anwendung neuer Alarm-Algorithmen

Im Rahmen dieser Studie werden neue Alarm-Algorithmen offline auf Basis des Referenzdatensatzes getestet. Dabei müssen die im Referenzdatensatz bereitgestellten Informationen so genutzt werden, dass neue Verfahren unter denselben Bedingungen operieren wie das während der Datenaufzeichnung verwendete Alarmsystem. Dies bedeutet zum einen, dass einem neuen Alarm-Algorithmus zu jedem Zeitpunkt nur die für die Alarmgebung relevanten Daten zugänglich gemacht werden, die bei einer Online-Anwendung verfügbar wären. Zum anderen werden die Kenntnisse über die eingestellten Alarmgrenzen und Alarm-Deaktivierungszeiträume so genutzt, dass die Gegebenheiten während der Datenaufzeichnung möglichst genau widerspiegelt werden.

Neue Alarm-Algorithmen werden als Alternative zu den einfachen Schwellwertalarmen vorgeschlagen. Arrhythmie-Alarme, ADV-Alarme und sonstige ernste Alarme des Infinity Monitoring Systems[®] bleiben unverändert bestehen. Auch das Konzept der vom Pflegepersonal einstellbaren Alarmgrenzen wird beibehalten. So besteht für den Kliniker weiterhin die Möglichkeit für jeden Vitalparameter Bereiche festzulegen, die seiner Einschätzung nach alarmierungswürdig sind. Neue Alarme werden generiert durch einen Vergleich des mittels einer robusten Regressionsmethode extrahierten zugrunde liegenden Signals mit den bestehenden Alarmgrenzen. Die für die Signal-extraktion verwendete robuste Regressionsmethode ist der Repeated Median (Siegel, 1982), welcher in einem gleitenden Fenster angewandt wird. Hierzu ist eine Fensterbreite festzulegen. In Kapitel 9 wird untersucht, für welche Vitalparameter eine Umstellung von einfachen Schwellwertalarmen auf signalextraktionsbasierte Schwellwertalarme von

Vorteil sein kann, welche Fensterbreite für die einzelnen Vitalparameter gewählt werden sollte und welche Zusatzregeln für die Alarmgebung möglicherweise nötig sind. Im Folgenden werden allgemeine Regeln vorgestellt, die beachtet werden müssen, um neue Alarme mit den Alarmen des derzeitigen verwendeten Alarmsystems vergleichen zu können. Diese Regeln beziehen sich auf die Reihenfolge von Alarmmeldungen und die Alarmgebung in Alarm-Deaktivierungszeiträumen.

8.1 Alarmreihenfolge

Bei den von der eData[®]-Software aus dem Netz abgerufenen Alarmmeldungen handelt es sich jeweils um die zeitlich zuerst aufgetretenen Alarme. Weitere Alarme, die parallel zu einem bereits aktiven Alarm ausgelöst werden, registriert die Aufzeichnungssoftware nicht. Ist z.B. ein einfacher Schwellwertalarm für den arteriellen systolischen Blutdruck mehrere Minuten aktiv (d.h. der Alarmgrund liegt während dieser Minuten die ganze Zeit vor), wird ein während dieser Zeit auftretender ADV-Alarm für die Sauerstoffsättigung zwar auf dem Monitorbildschirm angezeigt, aber nicht in der Alarmdatei abgespeichert.

Aus diesem Grund muss bei der Erzeugung neuer Alarme ebenfalls auf die Alarmreihenfolge geachtet werden. Werden z.B. neue Alarme für den systolischen und mittleren arteriellen Blutdruck generiert, muss jeder Alarm für eine Variable ignoriert werden, der während des noch aktiven Alarms der anderen Variablen auftritt.

8.2 Alarm-Deaktivierungszeiträume

Im Folgenden werden zwei Alarm-Deaktivierungstypen unterschieden. Zum einen kann die Alarmfunktion für einen bestimmten Vitalparameter gezielt deaktiviert werden. Zum anderen besteht durch Betätigung der „Alle Alarme Aus“-Taste die Möglichkeit die Alarmfunktion für *alle* Vitalparameter für bis zu drei Minuten auszuschalten.

Liegt der erste Fall vor, werden in diesem Zeitraum für die entsprechende Variable auch keine Alarme mittels eines neuen Alarm-Algorithmus gegeben.

Der zweite Fall gilt im Referenzdatensatz für 8% der Beobachtungsdauer (siehe Tabelle 7.1). Hier würde eine Unterdrückung neuer Alarme im Alarm-Deaktivierungszeitraum in bestimmten Fällen zu einem unfairen Alarmsystem-Vergleich führen. Nach einer näheren Untersuchung der im Referenzdatensatz vorliegenden Daten lassen sich prin-

ziell zwei Situationen unterscheiden, in denen die „Alle Alarmer Aus“-Taste betätigt wird. Zum einen kann die Betätigung *vorbeugend* erfolgen, z.B. vor Beginn einer Pflegemaßnahme, um durch die Manipulation verursachte Alarmer zu unterdrücken (siehe Abb. 8.1). In solchen Situationen ist es sinnvoll im Deaktivierungszeitraum keine neuen Alarmer zuzulassen.

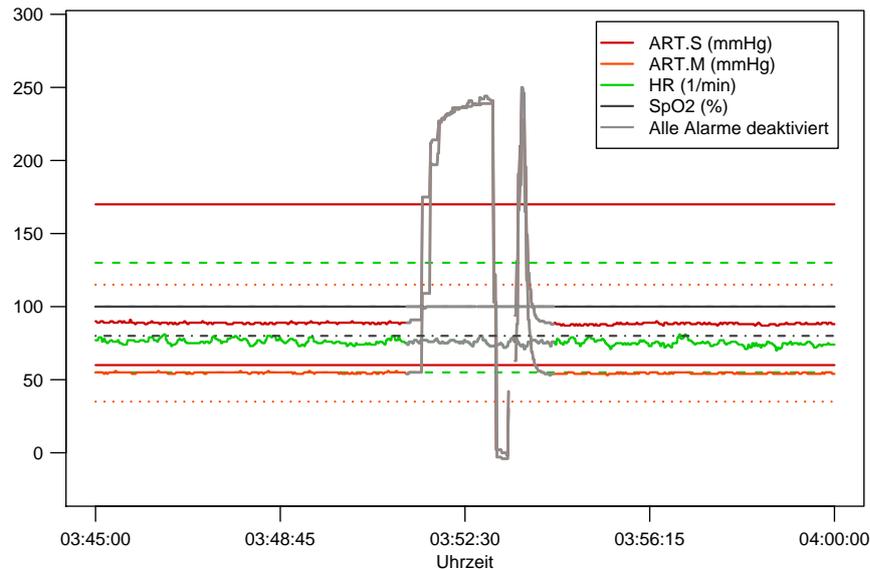


Abbildung 8.1: Beispiel für eine vorbeugende Alarmdeaktivierung

Zum anderen kommt es häufig vor, dass die „Alle Alarmer Aus“-Taste als *Reaktion* auf einen Alarm gedrückt wird, um diesen zu „quittieren“. Hier ist für die Generierung neuer Alarmer von Interesse, ob es sich bei dem Alarm um einen einfachen Schwellwertalarm handelt, der durch neue Alarmer „ersetzt“ werden soll (siehe Abb. 8.2), oder um einen Alarmtyp, der durch neue Alarm-Algorithmen nicht verändert wird (z.B. ADV-Alarmer).

Im ersten Fall treten Alarmer des auf Signalextraktion basierenden Systems für die entsprechende Situation möglicherweise verzögert auf und fallen in den Alarm-Deaktivierungszeitraum. Je nach klinischer Bedeutung der Situation würde durch Ignorieren des neuen Alarms entweder die Sensitivität des neuen Systems gegenüber dem alten Verfahren benachteiligt (falls Situation alarmrelevant) oder die Spezifität des neuen Verfahrens gegenüber den einfachen Schwellwertalarmen überschätzt (falls Situation nicht alarmrelevant). Daher werden in solchen Alarm-Deaktivierungszeiträumen neue Alarmer zugelassen.

Ist der Alarm des derzeit verwendeten Alarmsystems unmittelbar vor der Alarm-

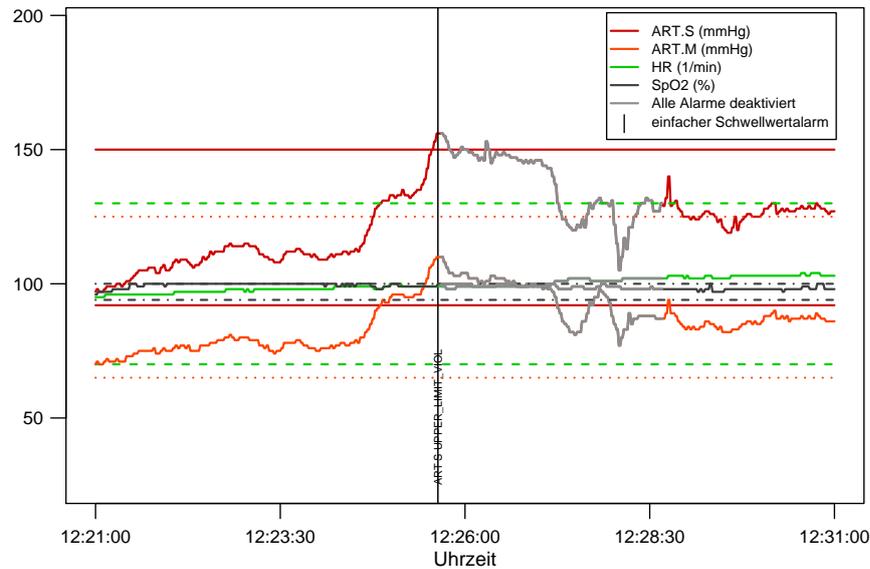


Abbildung 8.2: Beispiel für eine Alarm-Deaktivierung als Reaktion auf einen Schwellwertalarm

Deaktivierung dagegen von einem Typ, der vom neuen Alarm-Algorithmus nicht ersetzt wird, lösen Grenzüberschreitungen bzw. Grenzünterschreitungen des extrahierten Signals im Alarm-Deaktivierungszeitraum keine Alarme aus.

Für die Offline-Generierung neuer Alarme müssen Alarm-Deaktivierungen, die vorbeugend oder als Reaktion auf einen Alarm auftreten anhand der Daten unterschieden werden. Für diese Untersuchung wurde festgelegt, dass eine Alarm-Deaktivierung als Reaktion auf einen Alarm angesehen wird, falls der letzte Alarm des Infinity Monitoring Systems[®] zum Zeitpunkt der Alarm-Deaktivierung noch aktiv ist oder nicht mehr als 60 Sekunden vor Betätigung der „Alle Alarme Aus“-Taste ausgelöst wurde. Alle anderen Alarm-Deaktivierungen werden als vorbeugende Deaktivierungen betrachtet. In den für diese Analyse verwendeten Daten traten in 982 Beobachtungsstunden insgesamt 1630 Alarm-Deaktivierungen auf, von denen 72% nach der eben eingeführten Definition als Reaktion auf einen Alarm erfolgten und entsprechend 28% vorbeugend. In 54% der Fälle wurde die Alarmfunktion als Reaktion auf einen einfachen Schwellwertalarm deaktiviert.

Die Alarm-Deaktivierungsperioden führen zu einer weiteren Regel für die Generierung neuer Alarme. Diese rührt daher, dass beim Infinity Monitoring System[®] ein durch Betätigung der „Alle Alarme Aus“-Taste quittierter Alarm erneut ausgelöst wird, falls der Alarmgrund nach Ablauf der Deaktivierungsfrist immer noch besteht. Eine solche Situation findet sich z.B. in Abb. 8.3. Für die untersuchten Patientendaten kommen solche Fälle insgesamt 155 mal vor, d.h. bei 2.6% aller Alarme handelt es sich nicht um

neue Alarmsituationen, sondern um wiederholte Alarmierungen für eine bereits zuvor alarmierte Situation. Um einen fairen Vergleich der Systeme sicherstellen zu können, muss diese Regel auch für die Auslösung von Alarmen neuer Alarm-Algorithmen gelten.

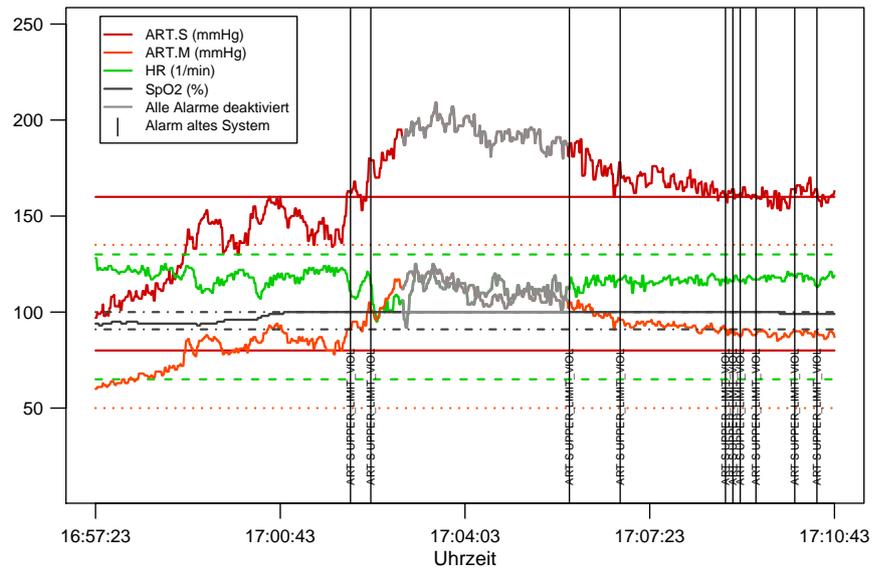


Abbildung 8.3: Beispiel für eine erneute Alarm-Auslösung nach Ablauf der Alarm-Deaktivierungsfrist, da der Alarmgrund immer noch besteht

Kapitel 9

Vergleich von Alarm-Algorithmen

Bevor ein umfassender Vergleich alter und neuer Alarm-Algorithmen stattfinden kann, wird zunächst überprüft, für welche Vitalparameter eine Umstellung von einfachen Schwellwertalarmen auf signalextraktionsbasierte Schwellwertalarme von Vorteil sein kann, welche Fensterbreite für die einzelnen Vitalparameter gewählt werden sollte und welche Zusatzregeln für die Alarmgebung möglicherweise sinnvoll und nötig sind. Dies geschieht anhand einer Teilstichprobe aus dem Referenzdatensatz. Für den endgültigen Vergleich der Alarm-Algorithmen werden die übrigen Daten verwendet.

9.1 Auswahl neuer Alarm-Algorithmen

Im vorliegenden Referenzdatensatz werden über 90% aller einfachen Schwellwertalarme durch den systolischen und mittleren arteriellen Blutdruck, die Sauerstoffsättigung und die Herzfrequenz ausgelöst. Auch die Atemfrequenz verursacht viele Schwellwertalarme, allerdings wurde diese nur in acht Fällen per Alarmsystem überwacht und es treten gemäß der Annotationen nur drei alarmrelevante Zustände für diese Variable auf. Daher konzentriert sich die folgende Untersuchung auf die Verbesserung der Alarmgebung für die drei zuerst genannten Vitalparameter. Ein gutes Abschneiden der neuen Verfahren für diese Variablen vorausgesetzt, könnten signalextraktionsbasierte Schwellwertalarme auch eine Alternative zu einfachen Schwellwertalarmen bei anderen Vitalparametern sein, z.B. bei pulmonalarteriellen Drücken, zentralvenösem Druck, Atemfrequenz oder Puls. Dies kann allerdings im Rahmen dieser Studie nicht überprüft werden, da nicht ausreichend Fälle vorhanden sind, bei denen diese Vitalparameter überwacht wurden.

Anhand eines Teils des Referenzdatensatzes werden im Folgenden verschiedene Signal-extraktionsmethoden in Kombination mit verschiedenen Alarmregeln auf ihre Eignung überprüft. Dazu werden zunächst so viele Fälle des Referenzdatensatzes zufällig ausgewählt bis eine Beobachtungsdauer von ungefähr einem Drittel der Beobachtungsdauer (ca. 330 Stunden) des gesamten Referenzdatensatzes erreicht ist. In Tabelle C.1 sind die so ausgewählten 23 Fälle aufgeführt. Die Beobachtungsdauer beträgt 334 Stunden. Tabelle C.1 zeigt außerdem für jeden Fall die Anzahl der pro Stunde durch die verschiedenen Vitalparameter ausgelösten einfachen Schwellwertalarme. Hier sind z.T. große Unterschiede zu beobachten, so dass es für die zukünftigen Untersuchungen sinnvoll erscheint, das Abschneiden eines neuen Alarm-Algorithmus nicht nur insgesamt sondern auch nach Fällen getrennt zu betrachten. Die Alarmfrequenzen in der ausgewählten Teilstichprobe werden in Abbildung 9.1 mittels Boxplots den Alarmfrequenzen in den verbleibenden Datensätzen gegenübergestellt, welche für den endgültigen Vergleich der Alarmsysteme verwendet werden sollen. Auf diese Weise soll untersucht werden, ob die beiden Gruppen vergleichbar sind. Die Mediane der Alarmfrequenzen in den beiden Gruppen unterscheiden sich für die drei Vitalparameter nur unwesentlich. Bei der Herzfrequenz fällt in der Teilstichprobe ein großer Ausreißer auf. Bei diesem Fall werden,

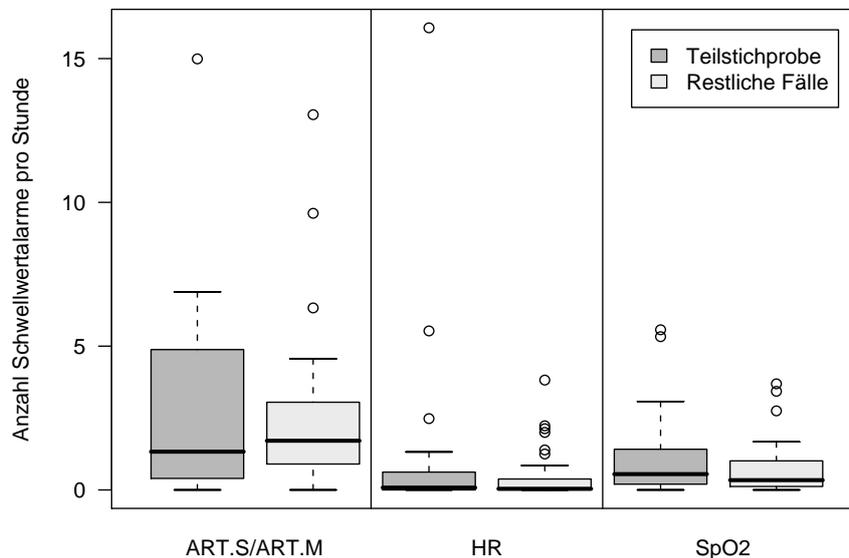


Abbildung 9.1: Vergleich der Anzahl einfacher Schwellwertalarme pro Stunde für die Teilstichprobe und die verbleibenden Fälle getrennt für arterielle Blutdrücke, Herzfrequenz und Sauerstoffsättigung

bedingt durch eine zu enge Einstellung der Alarmgrenzen, 16 HR-Schwellwertalarme pro Stunde ausgelöst. Für alle übrigen Fälle kommen pro Stunde weniger als sechs Schwellwertalarme für die Herzfrequenz vor. In der Gruppe der arteriellen Drücke treten ebenfalls vereinzelt sehr hohe Alarmfrequenzen auf, allerdings in beiden Gruppen. Tabelle C.1 gibt auch Auskunft über die Anzahl der für die einzelnen Vitalparameter aufgetretenen alarmrelevanten Situationen. Da bei der Entscheidung für einen neuen Alarm-Algorithmus auch die Sensitivität bezüglich der Erkennung alarmrelevanter Zustände berücksichtigt wird, muss gewährleistet sein, dass auch ausreichend viele dieser Zustände in der Teilstichprobe vorhanden sind. Die Teilstichprobe beinhaltet ungefähr ein Drittel der als alarmrelevant annotierten Schwellwertalarme für die arteriellen Blutdrücke. Neun alarmrelevante Situationen liegen in der Teilstichprobe für die Herzfrequenz vor. Da für diesen Vitalparameter insgesamt nur bei 30 Alarmen eine alarmrelevante Situation erkannt wurde, liegt der Anteil in der Teilstichprobe bei 30%. Für die Sauerstoffsättigung beinhalten die ausgewählten Fälle 50% aller alarmrelevanten Situationen.

Anhand der ausgewählten Fälle wird in den folgenden Abschnitten getrennt für die einzelnen Vitalparameter untersucht, welche neuen signalextraktionsbasierten Alarmregeln eine lohnenswerte Alternative zum einfachen Schwellwertsystem darstellen.

9.1.1 Arterielle Blutdrücke

Die Alarmgebung für den systolischen und mittleren arteriellen Blutdruck wird im Folgenden gemeinsam betrachtet. In einem ersten Schritt werden für beide Vitalparameter Repeated Median Signale für die Fensterbreiten 11, 21, ..., 121 Sekunden extrahiert. So erhält man ein breites Spektrum von Signalen: Eng an der Originalzeitreihe orientierte Signale für Fensterbreiten unter einer Minute und glattere Signale mit einer nicht zu großen Verzögerung beim Nachzeichnen von Niveauänderungen für Fensterbreiten bis zwei Minuten. Fehlen in einem Fenster einige Messwerte, wird die Schätzung auf Basis der übrigen Werte vorgenommen. Fehlen alle Werte im Fenster, ist der Signalschätzwert ebenfalls fehlend. Für die Alarm-Generierung wird die sogenannte Online-Schätzung, d.h. die Schätzung am Ende des jeweiligen Zeitfensters verwendet (siehe Abschnitt 2.2). Unter Beachtung der Regeln zu Alarm-Reihenfolge und Alarm-Deaktivierung (siehe Abschnitt 8) wird ein Alarm dann sofort ausgelöst, wenn das extrahierte Signal außerhalb der eingestellten Alarmgrenzen liegt. Abbildung 9.2 zeigt die Anzahl der durch dieses Vorgehen generierten Alarme im Vergleich zur Anzahl der einfachen Schwellwertalarme für den systolischen und mittleren arteriellen Blutdruck. Es fällt

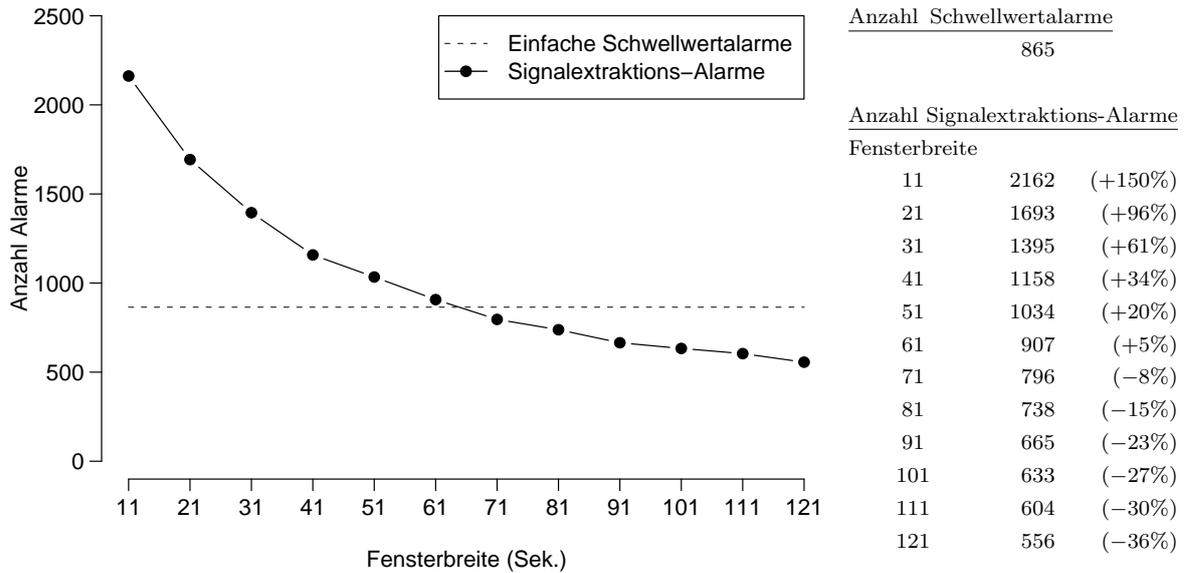


Abbildung 9.2: Anzahl neuer Alarme für den systolischen und mittleren arteriellen Blutdruck mittels Repeated Median Online-Signalextraktion für die Fensterbreiten 11, 21,..., 121 Sekunden im Vergleich zur Anzahl einfacher Schwellwertalarms

	Fensterbreite (Sek.)											
Neues System erzeugt	11	21	31	41	51	61	71	81	91	101	111	121
mehr Alarme	23	22	22	19	19	17	13	9	8	5	6	6
gleich viele Alarme	0	0	0	2	2	1	2	4	2	2	3	2
weniger Alarme	0	1	1	2	2	5	8	10	13	16	14	15

Tabelle 9.1: Anzahl Fälle, in denen der signalextraktionsbasierte Alarm-Algorithmus mehr, gleich viele und weniger Alarme als das einfache Schwellwertalarmsystem produziert

auf, dass der auf Signalextraktion basierende Alarm-Algorithmus für Fensterbreiten kleiner als 71 Sekunden mehr Alarme produziert als das Infinity Monitoring System[®]. Für Fensterbreite 11 ist die Alarmanzahl sogar 2.5-mal so hoch. Eine Erklärung für die zusätzlichen Alarme ist die Tatsache, dass vom einfachen Schwellwertalarmsystem eine Alarmbestätigungsfrist von einigen Sekunden verwendet wird. Im Falle des arteriellen Blutdrucks muss eine Grenzverletzung mindestens vier Sekunden andauern, bevor ein Alarm gegeben wird. Die bisher vorgestellten neuen Alarme werden bei einer Grenzverletzung des Signals sofort ausgelöst. Bei kleinen Fensterbreiten hat dies besonders große Auswirkungen, da das extrahierte Signal weniger glatt, d.h. noch sehr am Original-Signal orientiert ist.

Aber auch für Fensterbreiten, bei denen insgesamt eine deutliche Alarmreduktion erreicht wird (z.B. Fensterbreite 121), ist bei einzelnen Patienten die Alarmanzahl des

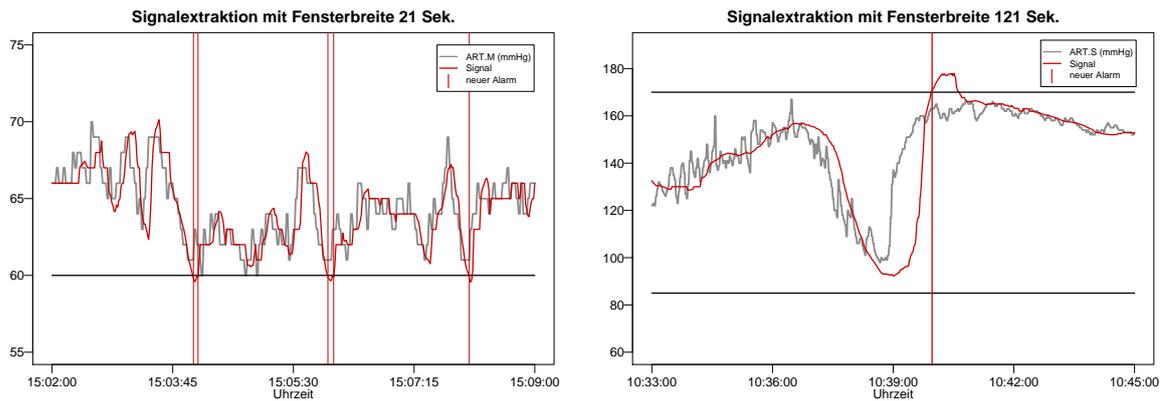


Abbildung 9.3: Beispiele für Unter- bzw. Überschätzungen des Signals (Fensterbreiten 21 und 121 Sek.), die zu zusätzlichen Alarmen führen

neuen Verfahrens dennoch höher (siehe Tabelle 9.1). Eine Ursache hierfür liegt darin, dass es bei der Repeated Median Schätzung für das Ende des Zeitfensters häufig zu Unter- und Überschätzungen des Signals kommt. In Abbildung 9.3 sind für die Fensterbreiten 21 und 121 Sekunden Beispiele von Unter- und Überschätzungen dargestellt, die zu unerwünschten zusätzlichen Alarmen führen.

Weiter verdeutlicht wird diese Problematik durch Abbildung C.1 im Anhang. Diese zeigt anhand von Boxplots jeweils für den systolischen und den mittleren arteriellen Blutdruck und verschiedene Fensterbreiten, wie hoch für die Fälle in der Teilstichprobe der Anteil der Fenster ist, in denen der geschätzte Signalwert größer als das Maximum bzw. kleiner als das Minimum der Werte im Fenster ist. Für kleine Fensterbreiten kann es bei bestimmten Patienten vorkommen, dass über 30% der geschätzten Signalwerte Über- oder Unterschätzungen sind. Das „rohe“ Online Repeated Median Signal sollte somit nicht als Basis für die Generierung von Alarmen verwendet werden, da durch die zu Unrecht ausgelösten Alarme ein falscher Eindruck über den Patientenzustand erweckt wird. Aus diesem Grund wird das Online-Signal im Folgenden so modifiziert, dass alle Signalwerte auf Werte zwischen Minimum und Maximum der Originalwerte im Fenster beschränkt werden. Bei einer Überschreitung des Maximums, wird der Signalwert auf das Maximum und bei einer Unterschreitung des Minimums entsprechend auf das Minimum gesetzt.

Abbildung 9.4 zeigt die Anzahlen der auf diesem modifizierten Signalextraktions-Algorithmus basierenden Alarme im Vergleich zur Anzahl einfacher Schwellwertalarme. Es zeigt sich, dass auch bei Beschränkung des Signals auf Werte zwischen Minimum und Maximum der Originalwerte im Fenster für kleinere Fensterbreiten mehr neue Alarme als einfache Schwellwertalarme generiert werden. Auch hier werden somit durch Verwendung einer Alarmbestätigungsfrist für Schwellwertalarme auf Basis der Ori-

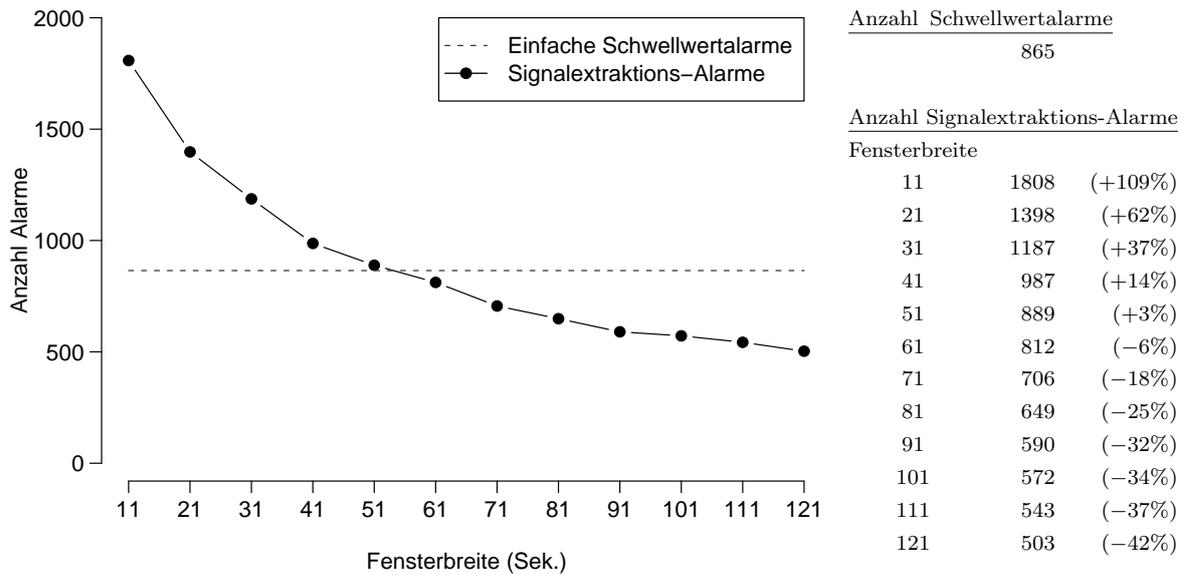


Abbildung 9.4: Anzahl neuer Alarme für den systolischen und mittleren arteriellen Blutdruck mittels Repeated Median Online-Signalextraktion mit Beschränkung auf Werte zwischen Minimum und Maximum der Originalwerte im Fenster für die Fensterbreiten 11, 21,..., 121 im Vergleich zur Anzahl einfacher Schwellwertalarms

	Fensterbreite (Sek.)											
Neues System erzeugt	11	21	31	41	51	61	71	81	91	101	111	121
mehr Alarme	23	22	21	18	14	13	8	4	2	1	2	1
gleich viele Alarme	0	0	0	1	5	3	3	3	4	5	4	3
weniger Alarme	0	1	2	4	4	7	12	16	17	17	17	19

Tabelle 9.2: Anzahl Fälle, in denen der signalextraktionsbasierte Alarm-Algorithmus mehr, gleich viele und weniger Alarme als das einfache Schwellwertalarmsystem produziert

nalwerte mehr Alarme gefiltert als durch eine robuste Signalextraktion mit kleiner Fensterbreite. Die generierten Alarme für Fensterbreiten ab 61 Sekunden liefern aber zufriedenstellende Alarmreduktionen. Allerdings sind Fensterbreiten größer als 91 Sekunden zu bevorzugen, da hier für die einzelnen Patienten nur in wenigen Ausnahmefällen mehr neue Alarme als einfache Schwellwertalarms auftreten (siehe Tabelle 9.2) und die Differenz der Alarmanzahlen in diesen Fällen auch nie mehr als zwei Alarme beträgt. Unterschiede dieser Größenordnung sind vernachlässigbar, da die Alarmanzahl, die bei einer echten Online-Anwendung eines neuen Alarm-Algorithmus resultieren würde, in dieser Untersuchung ohnehin nicht exakt feststellbar ist. Die hier angegebenen Anzahlen sind vermutlich eher etwas größer, als bei einer Online-Anwendung. Situationen, in denen möglicherweise mehr neue Alarme registriert werden, als bei der klinischen

Anwendung tatsächlich gegeben würden, sind Alarm-Deaktivierungszeiträume, die als Reaktion auf einen einfachen Schwellwertalarm für den systolischen oder mittleren arteriellen Blutdruck erfolgen und in denen alle neuen Alarme gezählt werden, um eine adäquate Bewertung zu gewährleisten (siehe Abschnitt 8.2). Weiterhin können bei einem abrupten starken Abfall des Blutdrucks „scheinbar“ zusätzliche neue Alarme auftreten. Der neue Alarm-Algorithmus meldet z.B. eine Schwellwertverletzung des Signals, während das alte Alarmsystem statt eines Schwellwertalarms möglicherweise die Alarmmeldung „Signal unterhalb des Messbereichs“ ausgibt.

Abbildung C.2 stellt für Fensterbreite 91 beispielhaft dar, wie sich die Verwendung der Signalextraktionsalarme auf die Alarmfrequenz bei den einzelnen Fällen auswirkt. Hier sind in einem Streudiagramm die Anzahl einfacher Schwellwertalarme pro Stunde gegen die Differenz zwischen den Alarmanzahlen pro Stunde für einfache und signalextraktionsbasierte Schwellwertalarme aufgetragen. Es zeigt sich, dass für Fälle, bei denen die Alarmfrequenz ohnehin niedrig ist, keine deutliche Verbesserung erzielt werden kann. Für Fälle mit hoher Alarmfrequenz (über fünf einfache Schwellwertalarme pro Stunde) ist eine Verwendung von signalextraktionsbasierten Alarmen lohnenswerter.

Sensitivität

Da die Alarmanzahlen allein keine Aussagekraft in Bezug auf die Güte neuer Alarm-Algorithmen haben, wird zusätzlich überprüft, in welchem Umfang eine Alarmreduktion auf Kosten der Sensitivität in Bezug auf die Erkennung alarmrelevanter Zustände geht. Die Sensitivität des einfachen Schwellwertsystems ist bedingt durch das Studiendesign 100%. Für die Signalextraktions-Alarme wird die Sensitivität gemäß der in Abschnitt 6.2.1 beschriebenen Vorgehensweise bestimmt. Ausgehend von den einfachen Schwellwertalarmen für den systolischen und mittleren arteriellen Blutdruck, die in den Annotationen die Bewertung alarmrelevant erhalten haben, werden alarmrelevante Zustände identifiziert. Um jeden dieser Schwellwertalarme wird ein Intervall der Länge ± 60 Sekunden gelegt. Ein alarmrelevanter Zustand wird durch den neuen Alarm-Algorithmus erkannt, wenn das neue Verfahren entweder innerhalb des Intervalls einen Alarm auslöst oder wenn ein neuer Alarm zwar vor Beginn des Intervalls anfängt, aber zum Auslösungszeitpunkt des alarmrelevant annotierten Schwellwertalarms noch aktiv ist. Die Sensitivität bestimmt sich als Anteil der vom neuen Alarm-Algorithmus erkannten an allen alarmrelevanten Situationen. Zum Vergleich werden nicht nur die Sensitivitäten für die Intervalllänge ± 60 Sekunden, sondern auch für kleinere (± 5 , ± 10 und ± 30 Sekunden) und größere (± 120 und ± 180 Sekunden) Längen berechnet (siehe Tabelle 9.3).

Im oberen Teil der Tabelle beziehen sich die angegebenen Sensitivitäten auf die Erken-

		Fensterbreite											
	Sek.	11	21	31	41	51	61	71	81	91	101	111	121
Insgesamt	±5	100	96.2	81.2	66.2	54.1	54.9	51.1	52.6	51.9	50.4	51.9	51.9
	±10	100	99.2	94.0	85.0	72.9	64.7	59.4	57.1	56.4	54.9	56.4	56.4
	±30	100	100	97.0	91.0	83.5	78.2	73.7	71.4	69.9	69.2	68.4	67.7
	±60	100	100	97.7	94.7	90.2	83.5	80.5	78.9	78.2	75.2	76.7	75.2
	±120	100	100	99.2	97.0	92.5	85.7	85.0	82.7	82.7	82.0	82.0	80.5
	±180	100	100	99.2	97.7	94.7	88.7	89.5	86.5	86.5	86.5	87.2	85.7
Technisch falsch	±5	100	98.6	79.5	57.5	46.6	47.9	46.6	47.9	43.8	41.1	42.5	42.5
	±10	100	100	94.5	82.2	69.9	60.3	52.1	52.1	50.7	45.2	46.6	46.6
	±30	100	100	97.3	87.7	76.7	67.1	64.4	63.0	58.9	56.2	56.2	57.5
	±60	100	100	98.6	93.2	84.9	72.6	71.2	69.9	67.1	64.4	64.4	64.4
	±120	100	100	100	97.3	89.0	76.7	76.7	74.0	74.0	72.6	71.2	71.2
	±180	100	100	100	98.6	93.2	82.2	83.6	79.5	79.5	79.5	79.5	79.5
Technisch richtig	±5	100	93.3	83.3	76.7	63.3	63.3	56.7	58.3	61.7	61.7	63.3	63.3
	±10	100	98.3	93.3	88.3	76.7	70.0	68.3	63.3	63.3	66.7	68.3	68.3
	±30	100	100	96.7	95.0	91.7	91.7	85.0	81.7	83.3	85.0	83.3	80.0
	±60	100	100	96.7	96.7	96.7	96.7	91.7	90.0	91.7	88.3	91.7	88.3
	±120	100	100	98.3	96.7	96.7	96.7	95.0	93.3	93.3	93.3	95.0	91.7
	±180	100	100	98.3	96.7	96.7	96.7	96.7	95.0	95.0	95.0	96.7	93.3

Tabelle 9.3: Sensitivitäten (in %) der auf dem Online-Signal (durch Minimum und Maximum der Werte im jeweiligen Fenster beschränkt) basierenden Alarm-Algorithmen für verschiedene Fensterbreiten und Intervalle um einen als alarmrelevant annotierten Schwellwertalarm.

nung aller alarmrelevanten Zustände. Insgesamt treten in den vorliegenden Daten 133 dieser Situationen auf (siehe Tabelle C.1). Es zeigt sich, dass der Signalextraktions-Alarm-Algorithmus für Fensterbreiten größer als 91 Sekunden – welche sowohl insgesamt als auch für die einzelnen Patienten eine lohnenswerte Alarmreduktion liefern – über eine Sensitivität von 78.2% nicht hinauskommt. Bei Fensterbreite 121, welche die größte Alarmreduktion liefert, beträgt die Sensitivität nur 75.2%. Selbst für Fensterbreiten, bei denen deutlich mehr neue Alarme als einfache Schwellwertalarme erzeugt werden, bleibt die Sensitivität häufig unter 100%. Eine Vergrößerung des alarmrelevanten Zeitraums auf ±3 Minuten um einen Schwellwertalarm mit Annotation alarmrelevant verbessert die Sensitivitäten um einige Prozentpunkte auf bis zu 87.2% für Fensterbreiten größer als 91 Sekunden. Diese Werte sind jedoch im Hinblick auf das im Studienplan formulierte Ziel der Nicht-Unterlegenheit in Bezug auf die Sensitivität mit einer Nicht-Unterlegenheitsschranke von 5% noch nicht zufriedenstellend.

Das schlechte Ergebnis kann teilweise dadurch erklärt werden, dass in den Annotatio-

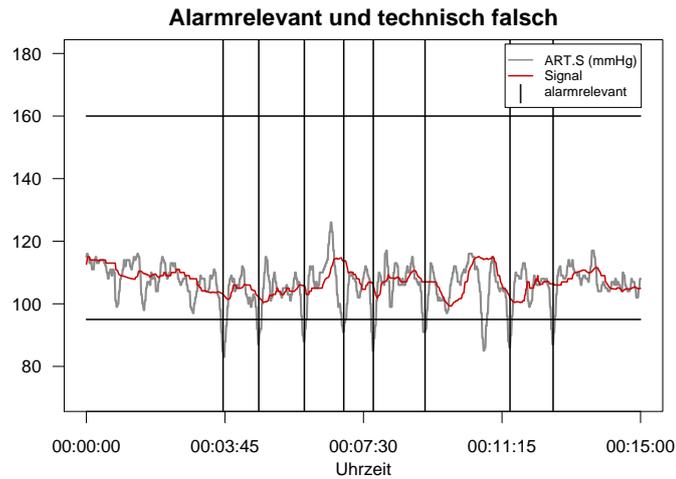


Abbildung 9.5: Beispiele für alarmrelevante Zustände aufgrund eines technischen Problems, welche vom neuen Alarm-Algorithmus (Fensterbreite 91) nicht erkannt werden

nen sowohl technisch richtige als auch technisch falsche Schwellwertalarme eine alarmrelevante Situation anzeigen können. Im Falle eines technisch falschen Alarms handelt es sich bei der alarmrelevanten Situation um ein länger andauerndes technisches Problem, was sich in der Originalzeitreihe häufig durch Ausreißer bemerkbar macht. Diese werden von der robusten Signalextraktionsmethode ignoriert und die entsprechenden alarmrelevanten Situationen werden nicht erkannt. Streng genommen ist in solchen Situationen zwar ein Alarm nötig, allerdings kein Schwellwertalarm sondern ein ADV-Alarm. In [Abbildung 9.5](#) sind Beispiele für alarmrelevante Situationen aufgrund eines technischen Problems dargestellt, die vom neuen Alarm-Algorithmus (für die Grafik wurde Fensterbreite 91 gewählt) nicht erkannt werden. Eine nähere Analyse für den Signalextraktions-Alarm-Algorithmus mit Fensterbreite 91 ergibt, dass von insgesamt 73 alarmrelevanten Zuständen technischer Ursache 24 nicht alarmiert werden. Auffällig ist, dass sämtliche dieser nicht erkannten Situationen bei demselben Patienten auftreten.

Die Sensitivitäten bei ausschließlicher Betrachtung der alarmrelevanten Situationen aufgrund eines technischen Problems finden sich im mittleren Teil von [Tabelle 9.3](#). Im Vergleich zur Gesamt-Sensitivität sind diese Sensitivitäten etwa zehn Prozentpunkte niedriger. Der untere Teil der Tabelle zeigt entsprechend die Sensitivitäten für den Fall, dass nur alarmrelevante Situationen betrachtet werden, bei denen der mit alarmrelevant annotierte Schwellwertalarm als technisch richtig bewertet wurde. Hierbei handelt es sich um insgesamt 60 Situationen. Die Sensitivitäten für diesen Fall sind deutlich besser. Werden wieder die Signalextraktionsalarme für Fensterbreite 91 näher untersucht, zeigt sich, dass von den alarmrelevanten Zuständen mit technisch korrekter Daten-

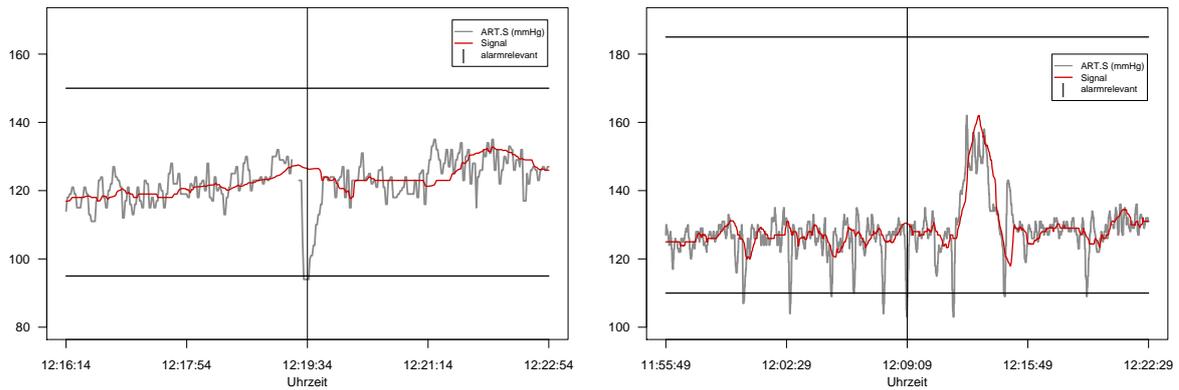


Abbildung 9.6: Alarmrelevante Zustände mit Annotation technisch richtig, welche vom neuen Alarm-Algorithmus (Fensterbreite 91) nicht erkannt werden

aufzeichnung insgesamt fünf nicht erkannt werden, was in einer Sensitivität von 91.7% resultiert. In drei dieser Fälle sind die Abstände des nächsten neuen Alarms zum alarmrelevant bewerteten Schwellwertalarm zwar größer als 60 Sekunden, aber kleiner als acht Minuten (90, 139, 440 Sekunden). Eine Vergrößerung des zulässigen Abstandes bei der Sensitivitätsberechnung führt somit zu Sensitivitäten, die die Nicht-Unterlegenheit des neuen Systems unterstützen. Die zwei übrigen Fälle, in denen zeitnah überhaupt kein neuer Alarm auftritt, sind in Abbildung 9.6 dargestellt. In beiden Fällen erfolgt die Grenzverletzung der Originalzeitreihe abrupt und ist nur von kurzer Dauer, so dass sie vom extrahierten Signal gefiltert wird. Da in beiden Fällen aus den Annotationen ersichtlich ist, dass auf den Alarm eine Reaktion erfolgte, ist es möglich, dass der Verlauf der Zeitreihe durch eine Intervention beeinflusst wurde. Der Vitalparameter nimmt dann wieder unbedenkliche Werte an, ehe das extrahierte Signal die alarmrelevante Veränderung feststellen kann. Eine derartige Störung des „natürlichen“ Zeitreihenverlaufs aufgrund einer Reaktion auf einen Alarm ist ein unvermeidlicher Nachteil der hier durchgeführten Offline-Validierung. Ein ähnliches Problem wurde schon in Abschnitt 8.2 diskutiert – dort im Zusammenhang mit Alarm-Deaktivierungen als Reaktion auf einen Alarm. Fensterbreite 111 liefert ebenfalls gute Sensitivitäten, wobei auch nur die in Abbildung 9.6 dargestellten Situationen überhaupt nicht erkannt werden.

Zusatzregeln

Für den untersuchten neuen Alarm-Algorithmus können Zusatzregeln konstruiert werden, die darauf abzielen, überflüssige Alarme weiter zu reduzieren. So kann, wie beim einfachen Schwellwertalarmsystem, eine Alarmbestätigungsfrist von vier Sekunden eingeführt werden, die Alarme durch sehr kurze Grenzverletzungen des extrahierten Signals unterdrückt. Abbildung C.3 und Tabelle C.2 zeigen, dass sich hier bereits für Fen-

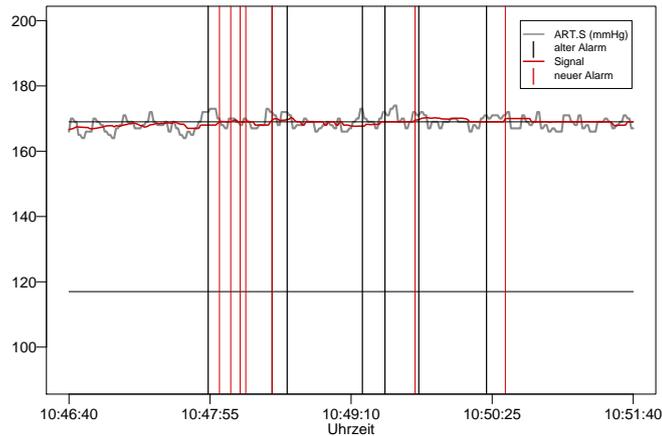


Abbildung 9.7: Beispiel für eine Situation in der viele neue Alarme aufgrund minimaler Schwankungen um die Alarmgrenze ausgelöst werden

sterbreiten ab 71 Sekunden hinsichtlich Alarmreduktion und Sensitivität gute Ergebnisse erzielen lassen. Verbesserungen gegenüber den Resultaten für die Alarm-Algorithmen ohne Alarmbestätigungsfrist werden aber nicht erreicht.

Abbildung 9.7 kann eine alternative Zusatzregel motivieren. Hier ist eine Situation dargestellt, in der neue Alarme aufgrund minimaler Schwankungen um die Alarmgrenze kurz hintereinander ausgelöst werden. Die Vermeidung solcher wiederholten Alarme durch Zusammenfügen zu einer langen Alarmsituation liefert Potential für eine deutliche Alarmreduktion ohne gleichzeitigen Sensitivitätsverlust. Eine Möglichkeit wäre die vorbehaltliche Verlängerung jedes Alarms für einige Sekunden und eine Fortführung des Alarms, wenn innerhalb dieser Zeit eine erneute Verletzung derselben Alarmgrenze erfolgt. Ein Nachteil dieser Methode ist, dass das Signal innerhalb der gesetzten Alarm-Verlängerungsfrist auch schon wieder deutlich innerhalb der Grenzen sein kann und ein noch aktiver Alarm in einem solchen Fall irreführend ist. Vielversprechender ist eine Regel, bei der ein Alarm so lange aktiv bleibt, bis das Signal wieder so weit innerhalb der Alarmgrenzen liegt, dass der Abstand zur vorher verletzten Alarmgrenze größer als ein bestimmter mmHg-Betrag. Bei Verwendung des extrahierten Signals für die Alarmgebung ist eine solche Regel besonders sinnvoll, wenn nicht gar erforderlich, da der Unterschied zwischen zwei aufeinanderfolgenden Werten u.U. auch nur ein Bruchteil eines mmHg betragen kann. Erfolgen solche Veränderungen in der Nähe einer Alarmgrenze, führt dies zu wiederholten Alarmen kurz hintereinander, wie in Abbildung 9.7 dargestellt. Ohne eine Regel zur Zusammenlegung solcher Situationen entsteht somit der unerwünschte Effekt, dass Alarme, die an anderen Stellen durch die robuste Glättung vermieden werden, in den soeben beschriebenen Situationen durch die Glättung unnötig

zusätzlich erzeugt werden. Die Zusatzregel kann natürlich auch zur Reduzierung von Schwellwertalarmen basierend auf der Originalzeitreihe angewendet werden, ist aber in Kombination mit einer vorherigen Glättung noch erfolgsversprechender. In den Tabellen C.3 und C.4 sowie Abbildung C.4 sind die Alarmhäufigkeiten und Sensitivitäten für den Fall zusammenfasst, dass ein Alarm erst endet, wenn das Signal wieder mehr als ein mmHg innerhalb der Grenzen ist. Hierdurch wird eine erhebliche Verbesserung der Alarmreduktion erreicht, wobei die Sensitivität nicht vermindert wird. Bei Wahl von Fensterbreite 91 wird die Alarmhäufigkeit bei einer Sensitivität von 91.7% (bei technisch richtiger Datenaufzeichnung) gegenüber den einfachen Schwellwertalarmen um 44% reduziert. Selbst Fensterbreite 61 liefert unter Verwendung der Zusatzregel 27% weniger Alarme und erreicht eine Sensitivität von 96.7%. Bei sechs Patienten werden zwar mehr Alarme ausgelöst, die Differenz der Alarmanzahlen ist aber vernachlässigbar gering (maximal zwei Alarme). Zum Vergleich wurde die Regel auch verwendet, um einfache Schwellwertalarme zusammenzufassen. Hier erreicht man eine Alarmreduktion von ca. 12%. Die vorherige Glättung der Zeitreihe ist somit von Vorteil. Durch eine Kombination der zuletzt vorgestellten Regel mit einer Alarmbestätigungsfrist von vier Sekunden können keine weiteren Verbesserungen der Ergebnisse erzielt werden. Für die signalextraktionsbasierten Blutdruck-Alarme werden daher im Folgenden keine Alarmbestätigungsfristen mehr betrachtet.

Signalschätzung für die Fenstermitte und adaptive Fensterbreitenwahl

Bisher wurden nur Alarme basierend auf der Online-Signalschätzung untersucht. Es besteht auch die Möglichkeit den Repeated Median Schätzwert für die Fenstermitte als Signalschätzung für das Ende des Zeitfensters zu verwenden. Diese Schätzung kann alternativ für die Erzeugung neuer Alarme benutzt werden. In Abbildung C.5 und Tabelle C.5 im Anhang finden sich die resultierenden Alarmhäufigkeiten und Sensitivitäten für die so generierten Alarme. Aufgrund der größeren Glattheit des Signals für die Fenstermitte werden schon ab Fensterbreite 31 Sek. weniger neue Alarme als einfache Schwellwertalarme erzeugt. Für Fensterbreite 51 wird eine Alarmreduktion von 35% bei einer Sensitivität (technisch richtig) von 91.7% erreicht. Allerdings erfolgt die Erkennung alarmrelevanter Zustände bei der Signalschätzung für die Fenstermitte nicht so zeitnah, was an den Sensitivitäten für kleinere Intervalle (± 5 bzw. ± 10 Sek.) um einen als „alarmrelevant“ annotierten Alarm zu erkennen ist. Die Sensitivitäten in Bezug auf alarmrelevante Situationen technischer Ursache und somit auch die Gesamt-Sensitivitäten sind außerdem wesentlich geringer als beim Online-Signal. Beides kann dadurch erklärt werden, dass die alternative Signalschätzung auf Niveauänderungen, Trendwechsel und andere Veränderungen des Original-Signals erst mit größerer

Verzögerung reagiert als das Online-Signal. Durch Anwendung der oben beschriebenen Zusatzregel zum Zusammenfügen aufeinanderfolgender Alarme aufgrund minimaler Schwankungen um die Alarmgrenze kann die Alarmhäufigkeit zwar weiter reduziert werden, die angemerkten Nachteile bleiben aber bestehen. Die Online-Schätzung ist somit als Basis für neue Alarme vorzuziehen.

Statt der Festlegung einer festen Fensterbreite für die Signalschätzung, besteht auch die Möglichkeit der *adaptiven* Fensterbreitenwahl abhängig von der Datensituation (Schettlinger, Fried und Gather, 2008). Die bereits gewonnenen Erkenntnisse über geeignete feste Fensterbreiten lassen sich für die Wahl der Parameter-Einstellungen der adaptiven Prozedur nutzen. So sollte z.B. der Wert für die minimal mögliche Fensterbreite nicht kleiner als eine Minute sein, da sonst zu viele zusätzliche Alarme erzeugt werden. Für die vorliegenden Daten wird für zwei Einstellungsmöglichkeiten getestet, ob sich durch Verwendung der adaptiv extrahierten Signale Verbesserungen der bisherigen Ergebnisse erzielen lassen. Für die minimale Fensterbreite werden dabei die Werte 60 und 90 Sekunden gewählt. Die maximale Fensterbreite wird in beiden Fällen auf 600 Sekunden festgelegt und es werden jeweils die letzten 30 Werte im Fenster zur Entscheidung über die Fenstergröße verwendet. Außerdem wird die Signalschätzung auf Werte zwischen Minimum und Maximum der Werte im jeweiligen Fenster beschränkt und für die Alarmgebung wird die oben beschriebene Zusatzregel angewandt. Relevante Verbesserungen der Alarmreduktion und Sensitivität lassen sich auf diese Weise aber nicht erzielen. Die Ergebnisse des adaptiven Verfahrens mit Wahl einer minimalen Fensterbreite von 60 Sekunden sind vergleichbar mit den Resultaten für die feste Fensterbreite 61 Sekunden. Die adaptive Prozedur mit minimaler Fensterbreite 90 schneidet ähnlich ab wie die Signalextraktion mit fester Fensterbreite 91. Um die Funktionsweise des neuen Alarm-Algorithmus so einfach wie möglich zu halten, konzentriert sich diese Vergleichstudie daher im Folgenden ausschließlich auf Alarm-Algorithmen basierend auf Signalextraktion mit fester Fensterbreite.

Falsch positive Alarme und PPV

Die Entscheidung darüber, welcher Alarm-Algorithmus für den endgültigen Vergleich verwendet wird, erfolgt nun anhand einer Betrachtung der Anzahl falsch positiver Alarme und Spezifitäten. Anhand der obigen Ergebnisse für Alarmhäufigkeiten und Sensitivitäten wird folgende Vorauswahl in Frage kommender Algorithmen vorgenommen: Die Auslösung eines Alarms basiert auf dem online extrahierten Repeated Median Signal, welches auf Werte zwischen Minimum und Maximum der Werte im Zeitfenster beschränkt ist. Zusätzlich gilt die Regel, dass der Alarm so lange aktiv ist, bis das extrahierte Signal wieder mehr als 1 mmHg innerhalb der Grenzen liegt. Geeignet er-

	altes	Signalextraktion mit Fensterbreite (Sek.)						
	System	61	71	81	91	101	111	121
falsch positiv	699	557 (-20%)	505 (-28%)	458 (-34%)	417 (-40%)	398 (-43%)	383 (-45%)	349 (-50%)
PPV	19%	12%	12%	13%	13%	13%	14%	15%

Tabelle 9.4: Anzahl Alarme außerhalb eines alarmrelevanten Zustandes (falsch positiv) und Positiver Prädiktiver Wert (PPV) = (Anz. Alarme - Anz. falsch positive Alarme)/Anz. Alarme

scheinen bisher die Fensterbreiten 61,...,121 Sekunden mit Alarmreduktionen zwischen 27% und 53% und Sensitivitäten bezüglich der Erkennung alarmrelevanter Zustände bei technisch richtiger Datenaufzeichnung zwischen 88.3% und 96.7%.

Gemäß Abschnitt 6.2.2 werden zur Bestimmung der falsch positiven Alarme und Spezifitäten nur die Alarme betrachtet, die außerhalb eines alarmrelevanten Zustandes auftreten und die auch während eines alarmrelevant annotierten Alarms nicht mehr aktiv sind. Es werden alle alarmrelevanten Zustände, d.h. unabhängig von technisch richtiger oder technisch falscher Datenaufzeichnung, berücksichtigt. In Tabelle 9.4 sind die so bestimmten Anzahlen unerwünschter Alarme für das einfache Schwellwertalarmsystem und die signalextraktionsbasierten Alarme mit Fensterbreiten zwischen 61 und 121 Sekunden zusammengefasst. Bei Verwendung der kleinsten Fensterbreite 61 wird die Häufigkeit falsch positiver Alarme nur um 20% reduziert. Die maximale Fensterbreite 121 liefert dagegen eine Verminderung der falsch positiven Alarme um 50%. Auf Basis der Information über die Anzahl falsch positiver Alarme kann auch der Positive Prädiktive Wert (PPV) der einzelnen Alarm-Algorithmen bestimmt werden, d.h. der jeweilige Anteil der Alarme, die während eines alarmrelevanten Zustands ausgelöst wurden. Für die einfachen Schwellwertalarme beträgt dieser Anteil 19%. Die neuen Alarm-Algorithmen erreichen dagegen je nach Fensterbreite nur Werte zwischen 12% und 15%. Diese Verschlechterung des Vorhersagewertes bedeutet jedoch nicht, dass die grundsätzliche Eignung der neuen Verfahren in Frage gestellt werden muss. Bei der Interpretation des Prädiktiven Wertes ist zu berücksichtigen, dass dieser durch folgenden Punkt beeinflusst wird: Die auf Signalextraktion basierenden Alarm-Algorithmen erzeugen anstatt vielen kurzen Alarmen für dieselbe Situation eher wenige längere Alarme. Der PPV setzt nun die Anzahl richtig positiver Alarme in Relation zur Anzahl aller Alarme. Durch die Tatsache, dass der Mediziner bei Verwendung eines neuen Alarm-Algorithmus somit für eine wesentlich geringere Anzahl Alarme über die Alarmrelevanz des Patientenzustandes entscheiden muss, ist die Verringerung des PPV weniger schwerwiegend und Konsequenz aus der gesunkenen Anzahl aller Alarme.

Spezifität

Für die Spezifitätsberechnung werden die zwei in Abschnitt 6.2.2 vorgestellten Ansätze verwendet und miteinander verglichen. Beim ersten Ansatz wird zunächst der gesamte nicht alarmrelevante Zeitraum in gleich große Intervalle eingeteilt. Für jedes Intervall und jedes Alarmsystem wird anschließend geprüft, ob ein Alarm auftritt. Die Spezifität eines Alarm-Algorithmus bestimmt sich als Anteil der Intervalle ohne Alarm an allen Intervallen. Zum Vergleich werden die unterschiedlichen Intervalllängen 60, 120, 180, 240, 300 und 600 Sekunden verwendet. Zeiträume, in denen alle Alarme durch Betätigung der „Alle Alarme Aus“-Taste deaktiviert waren, werden bei der Aufteilung berücksichtigt, da das neue Verfahren in manchen dieser Zeiträume Alarme auslösen kann. Zur Erreichung der anvisierten Intervalllänge werden aber nur die Sekunden mit aktivem Alarmstatus gezählt. Zeiträume, in denen sowohl für den systolischen als auch für den mittleren arteriellen Blutdruck die Alarmfunktion gezielt deaktiviert wurde, werden nicht berücksichtigt. Die vorgegebene Intervalllänge kann in manchen Situationen nicht exakt eingehalten werden, zum einen, wenn ein Zeitabschnitt zwischen zwei alarmrelevanten Zuständen kürzer als die vorgegebene Intervalllänge ist und zum anderen, wenn bei der Aufteilung eines nicht alarmrelevanten Zeitraums ein Rest verbleibt. Im letzteren Fall wird dieser Rest dem vorangegangenen Intervall zugeteilt, falls er kleiner als die Hälfte der vorgegebenen Intervalllänge ist.

Tabelle 9.5 zeigt für die gewählten Intervallbreiten die resultierenden Spezifitäten. Erwartungsgemäß sind die Spezifitäten umso höher, je kürzer die gewählte Intervalllänge ist. Für eine Länge von 60 Sekunden erhält man bereits Spezifitäten nahe 100%. Der *Zeitanteil*, in dem unerwünschte Alarme für den systolischen oder mittleren arteriellen Blutdruck auftreten, ist somit relativ gering. Betrachtet man den Anteil der Intervalle, in denen die Werte für den systolischen oder mittleren arteriellen Blutdruck zu irgendeinem Zeitpunkt außerhalb der Alarmgrenzen waren, ist dies aber einleuchtend (siehe Tabelle 9.6). Das einfache Schwellwertalarmsystem kann ausschließlich in solchen Intervallen einen Alarm geben. Deshalb sind die Zeitanteile mit unerwünschten Alarmen automatisch nach oben durch die in Tabelle 9.6 angegebenen Anteile beschränkt. Prinzipiell gilt dies auch für die auf Signalextraktion basierenden Algorithmen, obwohl diese in Ausnahmefällen wegen der Verzögerung auch in einem Intervall ohne Grenzverletzung einen Alarm auslösen können.

Für die einzelnen Zeitabschnittslängen sind die Spezifitätsunterschiede sowohl zwischen den verschiedenen neuen Alarm-Algorithmen als auch zwischen neuen Verfahren und einfachem Schwellwertalarmsystem sehr gering. Bei einer Intervalllänge von z.B. fünf Minuten sind die Spezifitäten der neuen Alarm-Algorithmen mit Werten zwischen

	Intervallbreite						
	60 Sek.	120 Sek.	180 Sek.	240 Sek.	300 Sek.	600 Sek.	
Altes System	97.0	95.1	93.2	91.8	90.6	85.0	
Neues System	61	97.4	95.3	93.6	92.1	90.8	85.6
mit Fenster-	71	97.6	95.7	93.9	92.5	91.3	86.1
breite	81	97.8	95.9	94.2	92.7	91.5	86.4
	91	98.0	96.2	94.6	93.2	92.0	87.1
	101	98.0	96.4	94.8	93.4	92.2	87.3
	111	98.1	96.5	95.0	93.7	92.5	87.7
	121	98.2	96.7	95.1	93.9	92.7	88.2

Tabelle 9.5: Spezifitäten (in %) auf Basis einer Aufteilung des nicht alarmrelevanten Zeitraums in gleich große Zeitintervalle (60, 120, 180, 240, 300, 600 Sek.) für das einfache Schwellwertalarmsystem und die auf Signalextraktion basierenden Alarm-Algorithmen

Intervallbreite Sek.	60	120	180	240	300	600
Anteil	7%	10%	13%	16%	18%	28%

Tabelle 9.6: Anteil der Intervalle, in denen die Originalwerte zu irgendeinem Zeitpunkt außerhalb der Alarmgrenzen liegen

90.8% und 92.7% trotz der beträchtlichen Unterschiede hinsichtlich der Alarmhäufigkeiten nur unwesentlich größer als die Spezifität des alten Schwellwertalarmsystems, welche 90.6% beträgt. Dies ist dadurch zu erklären, dass das einfache Schwellwertalarmsystem häufig eine Situation durch viele kurze Alarme anzeigt, während das auf Signalextraktion beruhende Verfahren eher wenige länger andauernde Alarme benötigt. Wie in Abschnitt 6.2.2 diskutiert, wird bei dem hier gewählten Intervall-basierten Ansatz aber nicht berücksichtigt, wie viele Alarme im Intervall ausgelöst werden, sondern nur ob ein Alarm auftritt. Am Beispiel des Vergleichs zwischen einfachem Schwellwertalarmsystem und Signalextraktionsalarmen mit Fensterbreite 91 ist in Tabelle 9.7 für Intervalllänge 240 Sekunden zusammengefasst, wie häufig welche Alarmanzahlen in den Intervallen vorkommen. Es ist wesentlich häufiger der Fall, dass in einem Intervall mehr alte als neue Alarme ausgelöst werden als umgekehrt.

Die Ergebnisse für den in Abschnitt 6.2.2 vorgestellten alternativen Ansatz zur Spezifitätsbestimmung, bei dem die Anzahl falsch positiver Alarme berücksichtigt wird, sind in Tabelle 9.8 angegeben. Während sich die Spezifität beim ersten Ansatz durch die Größe der Intervallbreiten beeinflussen lässt, ist hier die Wahl der „Worst case“-Anzahl FP_{max} entscheidend. Deswegen werden wieder verschiedene Möglichkeiten gegenübergestellt. Zwecks Vergleichbarkeit mit den Resultaten der ersten Methode wird FP_{max}

		Anzahl neuer Alarme									
		0	1	2	3	4	5	6	7	8	Σ
	0	4127	48	5	2	0	0	0	0	0	4182
	1	88	100	15	2	0	0	0	0	0	205
	2	21	43	18	5	0	1	1	0	0	89
Anzahl	3	6	22	12	1	0	0	0	0	0	41
alter	4	5	8	3	3	2	1	0	0	0	22
Alarme	5	1	7	2	2	0	0	0	0	0	12
	6	0	4	0	1	0	0	0	0	0	5
	7	0	0	1	0	0	0	0	0	0	1
	8	0	1	0	0	0	0	0	0	0	1
	Σ	4248	233	56	16	2	2	1	0	0	4558

Tabelle 9.7: Anzahl alter und neuer Alarme (Fensterbreite 91) in Intervallen der Länge 240 Sekunden

		„Worst-case“: Ein Alarm pro					
		60 Sek.	120 Sek.	180 Sek.	240 Sek.	300 Sek.	600 Sek.
Altes System		96.2	92.3	88.5	84.6	81.2	65.6
Neues System	61	96.9	93.9	90.8	87.7	84.8	72.2
mit Fenster-	71	97.2	94.4	91.7	88.9	86.1	73.7
breite	81	97.5	95.0	92.4	89.9	87.4	75.2
	91	97.7	95.4	93.1	90.8	88.5	77.4
	101	97.8	95.6	93.4	91.2	89.0	78.3
	111	97.9	95.8	93.7	91.6	89.5	78.9
	121	98.1	96.2	94.2	92.3	90.4	80.8

Tabelle 9.8: Spezifitäten (in %) auf Basis der Alarmanzahlen

dabei so bestimmt, dass man in nicht alarmrelevanten Zeiträumen schlimmstenfalls einen Alarm alle 60, 120, 180, 240, 300 oder 600 Sekunden erwartet. Auch hier erhält man in einigen Fällen recht hohe Spezifitäten, die Werte sind aber im Vergleich zum intervallbasierten Ansatz kleiner. Die Spezifitätsunterschiede zwischen alter und neuer Alarmgebungsmethode, aber auch zwischen den verschiedenen Fensterbreiten für die neuen Alarm-Algorithmen, kommen hier außerdem deutlicher zum Ausdruck, da die Anzahl Fehlalarme berücksichtigt wird.

Insgesamt ist bei den berechneten Spezifitäten noch zu bedenken, dass es sich bei 44% der falsch positiven einfachen Schwellwertalarme für die arteriellen Blutdrücke um herbeigeführte Alarme handelt. Obwohl für die neu erzeugten Alarme der genaue Anteil nicht feststellbar ist, werden auch hier viele ihre Ursache in einer Manipulation durch das medizinische Personal haben. Eine ausschließliche Berücksichtigung der

nicht herbeigeführten Alarme bei der Spezifitätsbestimmung würde somit zu höheren Spezifitätswerten führen, als sie hier ohnehin schon vorliegen.

Die Vergleiche der unterschiedlichen Spezifitätsdefinitionen liefern nützliche Informationen. Für den endgültigen Vergleich muss schließlich eine einzige Definition gewählt werden. Da der auf Intervall-Einteilung basierende Ansatz für die vorliegende Datensituation zu viele Nachteile aufweist, wird der zweite Ansatz bevorzugt. Hierfür ist noch ein geeigneter Sekundenwert für die Bestimmung der „Worst-case“-Anzahl auszuwählen. Hierbei erfolgt eine Orientierung an den im Referenzdatensatz auftretenden Anzahlen an Schwellwertalarmen für ART.S und ART.M pro Stunde für die einzelnen Patienten. Die höchste Alarmanzahl pro Stunde bei einem Patienten beträgt 15. Dies würde dem Auftreten eines Alarms alle vier Minuten entsprechen. Für die Spezifitätsberechnung bei Alarmen für invasive Blutdrücke wird somit FP_{max} so bestimmt, dass ein potentieller Alarm je 240 Sekunden nicht alarmrelevantem Zeitraum gezählt wird.

Auswahl eines Alarm-Algorithmus für die Validierung

Um zu einer Entscheidung für eine bestimmte Fensterbreite zu gelangen, müssen alle bisherigen Erkenntnisse in Betracht gezogen werden. Für die kleinste zur Auswahl stehende Fensterbreite 61 spricht die gute Sensitivität. Diese liegt als Einzige über der für die Nicht-Unterlegenheit festgelegten unteren Schranke von 95%. Die Alarmreduktion insgesamt ist mit 27% zufriedenstellend. Damit sich die Verringerung der Schwellwertalarme für die arteriellen Blutdrücke aber auch spürbar auf die Gesamt-Alarmfrequenz auswirkt (Schwellwertalarme für ART.S und ART.M machen 40% aller Alarme aus), sollte die Alarmreduktion jedoch idealerweise noch größer sein. Zudem beträgt die Alarmreduktion für die falsch positiven Alarme bei Fensterbreite 61 auch nur 20%. Der Algorithmus mit der größten betrachteten Fensterbreite 121 ist dagegen zwar hinsichtlich der Häufigkeit falsch positiver Alarme deutlich besser als das einfache Schwellwertsystem, die Sensitivität ist aber gleichzeitig auch deutlich schlechter als die Zielvorgabe. Für die Validierung wird daher als Kompromiss der Alarm-Algorithmus mit Fensterbreite 91 verwendet. Dieser liefert eine lohnenswerte Verminderung der Alarmhäufigkeit (44% insgesamt und 40% hinsichtlich der falsch positiven Alarme) und eine gute Sensitivität von 91.7%.

Vergleich nach Charbonnier und Gentil (2007)

Zum Abschluss der Auswertung für den betrachteten Teil des Referenzdatensatzes erfolgt für den ausgewählten Alarm-Algorithmus eine Anwendung der in Charbonnier und Gentil (2007) beschriebenen Auswertungsmethode, deren Prinzip in Abschnitt 6.3 erläutert wird (siehe Tabelle C.6). Bei Verwendung von Fensterbreite 91 zeigt sich, dass

bei 712 der 865 einfachen Schwellwertalarme gleichzeitig ein neuer Alarm auftritt, d.h. es existiert in diesen Fällen eine Überlappung der Alarmperioden (± 30 Sek.). Dagegen liegt bei 153 einfachen Schwellwertalarmen gleichzeitig kein neuer Alarm vor, d.h. diese Alarme werden vom neuen Verfahren unterdrückt. Eine deskriptive Auswertung der Alarmdauern dieser unterdrückten Alarme zeigt, dass es sich um Alarme handelt, die im Mittel nur ca. 6 und maximal 21 Sekunden andauern. Ein Blick in die Annotationen gibt Aufschluss darüber, ob die unterdrückten Alarme bedeutsam sind: Fünf der Alarme tragen die Annotation „alarmrelevant und technisch richtig“ und werden somit zu Unrecht unterdrückt. Dieses Resultat stimmt mit dem bei der Sensitivitätsbestimmung erhaltenen Ergebnis für das neue Verfahren überein (siehe Tabelle C.4). Beim umgekehrten Vergleich zeigt sich, dass bei 374 der neuen Alarme gleichzeitig mindestens ein alter Alarm vorliegt. Die Verzögerung (Delay) der neuen Alarme gegenüber den einfachen Schwellwertalarmen beträgt im Mittel nur 3.75 Sekunden. In 25% der Fälle erfolgt die Alarmgebung sogar früher, allerdings kommen auch Verzögerungen von bis zu 250 Sekunden vor. Bei 107 neuen Alarmen tritt gleichzeitig kein alter Alarm auf. Eine Untersuchung der einzelnen Situationen ergibt folgende Ursachen für die zusätzlichen Alarme: In vielen Fällen ist die Originalzeitreihe zwar außerhalb der Grenzen, aber nie länger als die eingestellte Alarmbestätigungsfrist. Ein Alarm auf Basis des extrahierten Signals kann aber durchaus auftreten. In anderen Fällen ist das extrahierte Signal nach Aufhebung einer Alarm-Deaktivierung noch außerhalb der Grenzen, die Originalzeitreihe aber nicht mehr. Sehr häufig kommt dies bei Blutabnahmen vor. Solche Alarme könnten durch eine Verlängerung der Alarmdeaktivierungsfrist (z.B. von drei auf vier Minuten) verhindert werden. Wie bereits erwähnt, kann es in manchen Situationen auch vorkommen, dass das Infinity System[®] statt eines Schwellwertalarms z.B. einen Alarm mit der Meldung „Werte unterhalb/oberhalb des Messbereichs“ gibt. Bei einer tatsächlichen Implementierung von signalextraktionsbasierten Alarmen sollte daher auf die Abstimmung mit anderen Alarmarten geachtet werden.

9.1.2 Herzfrequenz

Im Folgenden werden auf Signalextraktion basierende Alarm-Algorithmen für den Vitalparameter Herzfrequenz näher untersucht. Auch für diese Variable werden zunächst Repeated Median Signale für die Fensterbreiten 11, 21, ..., 121 extrahiert, wobei die Schätzung für das Ende des Zeitfensters erfolgt. Da genau wie bei den arteriellen Blutdrücken bei Verwendung des reinen Online-Signals Über- und Unterschätzungen auftreten, werden die Signalwerte zusätzlich auf Werte zwischen Minimum und Maxi-

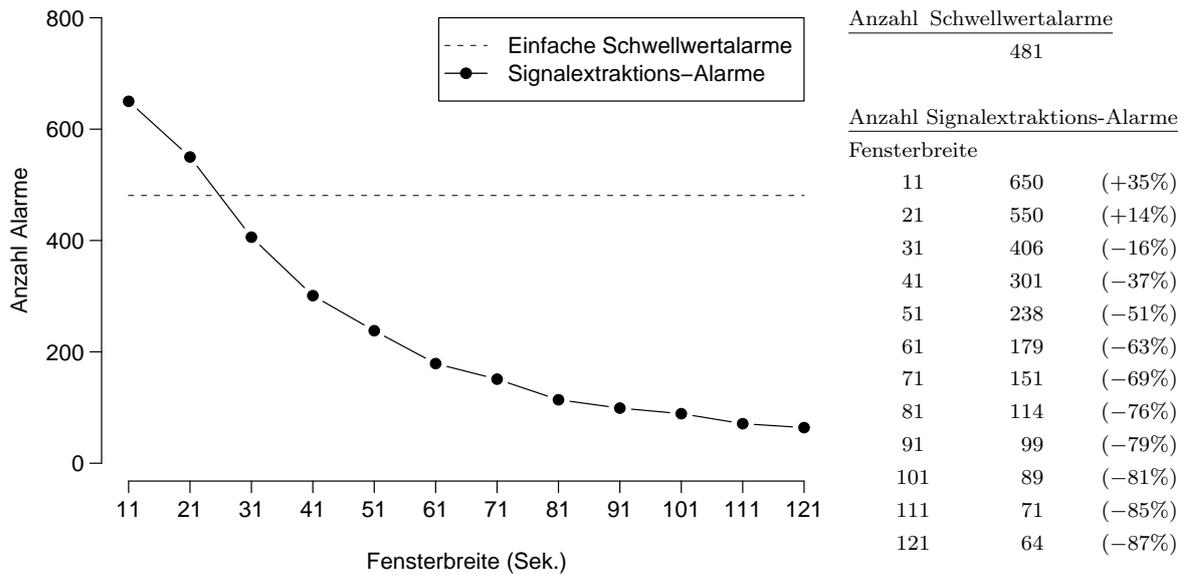


Abbildung 9.8: Anzahl neuer Alarme für die Herzfrequenz für die Fensterbreiten 11, 21,..., 121 Sekunden im Vergleich zur Anzahl einfacher Schwellwertalarms

Neues System erzeugt	Fensterbreite (Sek.)											
	11	21	31	41	51	61	71	81	91	101	111	121
mehr Alarme	6	4	1	0	0	0	0	0	0	0	0	0
gleich viele Alarme	13	14	14	11	12	11	11	11	11	11	11	11
weniger Alarme	4	5	8	12	11	12	12	12	12	12	12	12

Tabelle 9.9: Anzahl Fälle, in denen der neue Alarm-Algorithmus für die Herzfrequenz mehr, gleich viele und weniger Alarme als das einfache Schwellwertalarmsystem produziert

zum Originalwert im Fenster beschränkt. Zudem wird wieder die Regel verwendet, dass ein Alarm erst dann endet, wenn das extrahierte Signal wieder mehr als eine Einheit (hier: ein Herzschlag/min) innerhalb der Alarmgrenzen liegt. Zum Vergleich ist hier anzumerken, dass die Verwendung dieser Regel für Alarme basierend auf der Originalzeitreihe eine Alarmreduktion von 6% bewirkt. [Abbildung 9.8](#) zeigt die Anzahl der erzeugten neuen Alarme im Vergleich zur Anzahl einfacher Schwellwertalarms für die Herzfrequenz. Hier treten nur für die beiden kleinsten Fensterbreiten mehr neue Alarme als einfache Schwellwertalarms auf. Ab Fensterbreite 31 Sekunden werden weniger neue Alarme generiert. Bei einer Verwendung von Fensterbreite 51 beträgt die Alarmreduktion schon über 50%. Zu beachten ist in diesem Zusammenhang, dass bei der Herzfrequenz für Schwellwertalarms kürzere Alarmbestätigungsfristen als bei den arteriellen Blutdrücken gelten. Bei Überschreitung der oberen Alarmgrenze erfolgt bereits nach zwei Sekunden ein Alarm, bei Unterschreitung der unteren Grenze sofort.

		Fensterbreite											
	Sek.	11	21	31	41	51	61	71	81	91	101	111	121
Insgesamt	± 5	100	77.8	66.7	66.7	44.4	33.3	22.2	22.2	22.2	22.2	22.2	22.2
	± 10	100	100	66.7	66.7	66.7	55.6	44.4	33.3	22.2	22.2	22.2	22.2
	± 30	100	100	100	88.9	88.9	88.9	66.7	66.7	77.8	77.8	55.6	33.3
	± 60	100	100	100	100	88.9	88.9	88.9	88.9	88.9	88.9	88.9	88.9
	± 120	100	100	100	100	88.9	88.9	88.9	88.9	88.9	88.9	88.9	88.9
	± 180	100	100	100	100	88.9	88.9	88.9	88.9	88.9	88.9	88.9	88.9

Tabelle 9.10: Sensitivitäten (in %) der signalextraktionsbasierten Alarm-Algorithmen für die Herzfrequenz für verschiedene Fensterbreiten und Intervalle um einen als alarmrelevant annotierten Schwellwertalarm

Aus Tabelle 9.9 ist ersichtlich, dass für diejenigen Fensterbreiten, welche insgesamt eine Verringerung der Alarmanzahl liefern, nur bei Fensterbreite 31 für einen Patienten mehr Alarme vom neuen Alarm-Algorithmus ausgelöst werden. In diesem konkreten Fall handelt es sich aber nur um einen einzigen zusätzlichen Alarm. Für viele Patienten bleibt die Anzahl Alarme unverändert. Bei den meisten dieser Fälle kann die Alarmanzahl aber auch nicht weiter reduziert werden, da kein einfacher Schwellwertalarm für die Herzfrequenz vorkommt. Dagegen entfallen allein 376 der 481 einfachen Schwellwertalarme auf einen einzelnen Patienten. Diese hohe Alarmhäufigkeit bei einer Beobachtungsdauer von knapp 24 Stunden ist auf eine sehr enge Einstellung der Alarmgrenzen zurückzuführen und stellt eine große Ausnahme dar. Betrachtet man die Alarmreduktionen ohne den betreffenden Patienten, verbessern sich die Ergebnisse für einige Fensterbreiten deutlich, z.B. bei Fensterbreite 31 von 16% auf 25% Alarmreduktion.

Bei neun der insgesamt 481 einfachen Schwellwertalarme für die Herzfrequenz wurde das Vorliegen eines alarmrelevanten Zustandes annotiert, wobei sich diese neun Situationen auf nur insgesamt vier Patienten beziehen. In allen alarmrelevanten Situationen wurde der entsprechende Schwellwertalarm als technisch richtig bewertet. Für die Herzfrequenz wird ein alarmrelevanter Zustand als vom neuen System erkannt angesehen, wenn in einem Intervall von ± 30 Sekunden um den alarmrelevant annotierten Schwellwertalarm ein neuer Alarm auftritt oder wenn zum Zeitpunkt des Schwellwertalarms noch ein neuer Alarm aktiv ist. Das Intervall wird kleiner als bei den arteriellen Blutdrücken gewählt, weil bei der Herzfrequenz aus medizinischer Sicht kritische Situationen schneller erkannt werden müssen. In Tabelle 9.10 werden zum Vergleich auch wieder Sensitivitäten für kleinere und größere Intervalle angegeben. Für die Sensitivität des einfachen Schwellwertsystems muss bedingt durch das Studiendesign wieder

	altes System	Signalextraktion mit Fensterbreite (Sek.)			
		31	41	51	61
falsch positiv	470	397 (-16%)	293 (-38%)	230 (-51%)	171 (-64%)
PPV	2%	2%	3%	3%	4%

Tabelle 9.11: Anzahl Alarme außerhalb eines alarmrelevanten Zustandes (falsch positiv) und Positiver Prädiktiver Wert (PPV)

ein Wert von 100% angenommen werden. Bei Verwendung von Fensterbreite 31 erhält man ebenfalls eine Sensitivität von 100%. Für die Fensterbreiten 41 bis 61 werden acht der neun alarmrelevanten Situationen rechtzeitig erkannt. Bei Fensterbreite 41 erfolgt bei der verpassten Situation ein neuer Alarm mit einer Verzögerung von 44 Sekunden jedoch noch recht zeitnah.

Durch Verwendung einer Alarmbestätigungsfrist von zwei Sekunden für Verletzungen der oberen Alarmgrenze werden keine wesentlichen Verbesserungen der Ergebnisse erzielt. Dasselbe gilt für Alarme auf Basis einer Signalschätzung für die Fenstermitte und auf Basis adaptiver Fensterbreitenwahl (mit minimaler Fensterbreite 30, maximaler Fensterbreite 600 und Entscheidung über Veränderung der Fenstergröße anhand der letzten 15 Werte im Fenster). Diese Ansätze werden daher hier nicht weiter verfolgt. Bezüglich der Anzahl falsch positiver Alarme und Spezifitäten werden im Folgenden nur die Fensterbreiten 31 bis 61 betrachtet, da nur diese sowohl hinsichtlich Alarmreduktion und Sensitivität geeignet erscheinen. Tabelle 9.11 zeigt für die falsch positiven Alarme ähnliche Alarmreduktionen durch die neuen Algorithmen wie für alle Alarme insgesamt. Dies liegt daran, dass nur ein sehr geringer Anteil der Alarme richtig positiv ist. Die extrem kleinen Positiven Prädiktiven Werte resultieren aus der geringen Zahl von nur neun alarmrelevanten Situationen in über 300 Stunden Beobachtungsdauer. Für die Bestimmung der Spezifität auf Basis der Alarmanzahl wird angenommen, dass schlimmstenfalls ein potentieller falscher Alarm pro zehn Minuten Beobachtungsdauer auftritt. Dieser Wert kann mit den im Referenzdatensatz vorkommenden HR-Schwellwertalarmen pro Stunde für die einzelnen Patienten begründet werden. Abgesehen von dem bereits erwähnten Ausreißer (Fall 64) treten in einer Stunde bei allen Fällen höchstens sechs Schwellwertalarme pro Stunde auf. Dies würde einem Alarm alle zehn Minuten entsprechen. In Tabelle 9.12 finden sich die so bestimmten Spezifitäten. Bei der Entscheidung welche Fensterbreite für die Validierung verwendet soll, ist zu berücksichtigen, dass aufgrund der sehr niedrigen Anzahl alarmrelevanter Situationen das Verhalten des Algorithmus nicht zuverlässig eingeschätzt werden kann. Besonders

	altes System				
	Fensterbreite (Sek.)				
	31	41	51	61	
Spezifität	87.8	89.2	90.3	91.0	91.6

Tabelle 9.12: Spezifitäten (in %) für Herzfrequenzalarme

bei der Herzfrequenz ist aber eine zeitnahe Erkennung kritischer Situationen unabdingbar. Obwohl größere Fensterbreiten lohnenswertere Alarmreduktionen liefern, wird daher für die Validierung der Alarm-Algorithmus mit Fensterbreite 31 gewählt, da dieser alle zehn alarmrelevanten Situationen erkennt. Analysiert man die Alarme für diesen Alarm-Algorithmus nach der Methode von Charbonnier und Gentil (2007) erhält man die in Tabelle C.7 zusammengefassten Ergebnisse. Für keinen der 43 unterdrückten alten Alarme wurde ein alarmrelevanter Zustand annotiert. Dieses Ergebnis passt zu dem zuvor bestimmten Sensitivitätswert. Die 63 zusätzlichen neuen Alarme beziehen sich fast ausschließlich auf den bereits vorher erwähnten Patienten mit der extrem hohen Alarmfrequenz. Eine Hauptursache sind hier ebenfalls Situationen, in denen zwar sowohl die Originalzeitreihe als auch das extrahierte Signal über der Grenze liegen, die Originalzeitreihe aber für kürzere Zeit als die eingestellte Alarmbestätigungsfrist.

9.1.3 Sauerstoffsättigung

Die Sauerstoffsättigung (SpO₂) unterscheidet sich in einigen Punkten von den Vitalparametern Blutdruck und Herzfrequenz. Sie weist zum einen die Besonderheit auf, dass keine Werte über 100% angenommen werden können. Diese Bedingung muss ein für die alternative Alarmgebung extrahiertes Signal ebenfalls erfüllen, was aber leicht durch eine Beschränkung der Signalwerte auf Werte zwischen Minimum und Maximum der Originalwerte in dem für die Signalbestimmung verwendeten Fenster zu erreichen ist. Zeitreihen für die Sauerstoffsättigung weisen typischerweise eine eher geringe Variabilität auf, so dass durch Signalextraktion oft keine weitere Glättung möglich ist. Allerdings ist bei der Sauerstoffsättigung, bedingt durch die Messung mittels eines Clips am Finger, häufiger mit Ausreißern (z.B. Bewegungsartefakten) zu rechnen, die durch robuste Signalextraktion gefiltert werden können. Viele der Ausreißer führen beim Infinity Monitoring System[®] allerdings gar nicht zu SpO₂-Schwellwertalarmen, da eine vergleichsweise lange Alarmbestätigungsfrist von zehn Sekunden eingestellt ist. Die angeführten Punkte lassen eine Verwendung von signalextraktionsbasierten Alarmen auf den ersten Blick nur bedingt geeignet zur Reduzierung überflüssiger SpO₂-

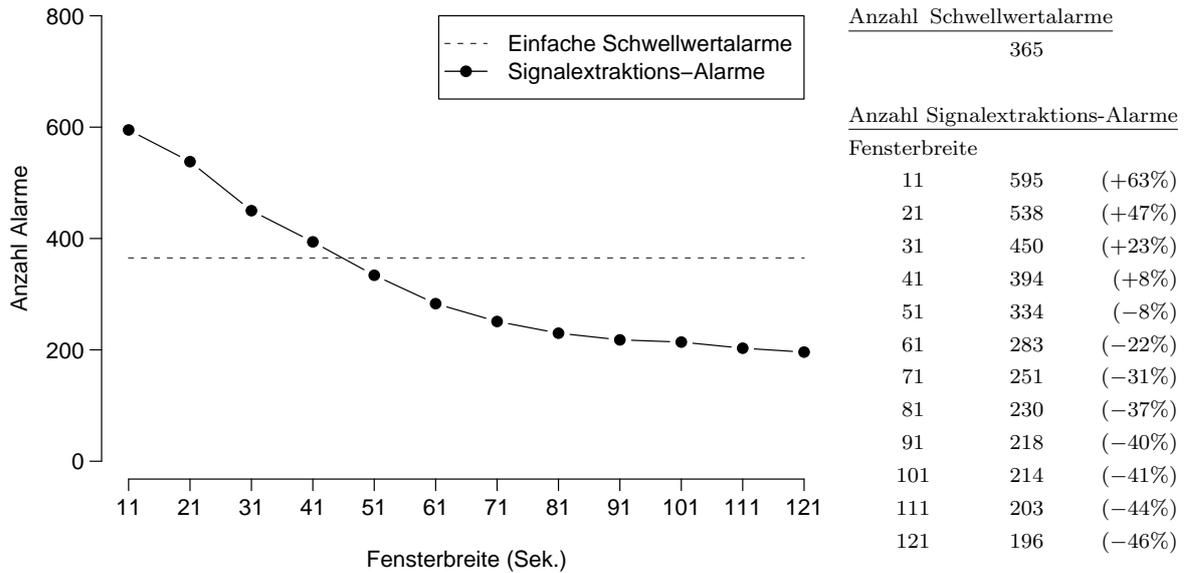


Abbildung 9.9: Anzahl neuer Alarme für die Sauerstoffsättigung für die Fensterbreiten 11, 21,..., 121 Sekunden im Vergleich zur Anzahl einfacher Schwellwertalarms

Neues System erzeugt	Fensterbreite (Sek.)											
	11	21	31	41	51	61	71	81	91	101	111	121
mehr Alarme	16	16	14	12	11	9	6	5	5	4	4	3
gleich viele Alarme	2	2	3	2	2	2	4	2	2	5	3	3
weniger Alarme	5	5	6	9	10	12	13	16	16	14	16	17

Tabelle 9.13: Anzahl Fälle, in denen der neue Alarm-Algorithmus für die Sauerstoffsättigung mehr, gleich viele und weniger Alarme als das einfache Schwellwertalarmsystem produziert

Schwellwertalarms bei möglichst unverminderter Sensitivität erscheinen. Dennoch wird die Eignung des neuen Verfahrens im Folgenden kurz untersucht. Auch in diesem Fall werden Repeated Median Signale für die Fensterbreiten 11, 21,..., 121 extrahiert, wobei sich die Schätzung auf das Ende des Zeitfensters bezieht. Die Signalwerte werden dabei auf Werte zwischen dem Minimum und dem Maximum der Originalwerte im Fenster beschränkt. Die Regel, dass ein Alarm erst endet, wenn das extrahierte Signal wieder mehr als eine Einheit (hier: ein Prozentpunkt) über der Alarmgrenze ist, wird ebenfalls verwendet. Zum Vergleich ist anzumerken, dass die Verwendung dieser Regel für Alarme basierend auf der Originalzeitreihe eine Alarmreduktion von 16% bewirkt. Abbildung 9.9 zeigt die Anzahl der offline generierten neuen Alarme im Vergleich zur Anzahl der einfachen Schwellwertalarms für die Sauerstoffsättigung. Erst ab einer Fensterbreite von 51 Sekunden werden vom neuen Alarm-Algorithmus weniger Alarme als vom einfachen Schwellwertsystem ausgelöst. Aber auch für einige von diesen Fen-

		Fensterbreite											
	Sek.	11	21	31	41	51	61	71	81	91	101	111	121
Insgesamt	± 5	100	100	97.1	68.4	61.0	56.6	50.0	50.0	48.5	47.1	49.3	50.0
	± 10	100	100	98.5	90.4	70.6	64.0	57.4	55.9	50.0	50.7	52.9	52.2
	± 30	100	100	98.5	94.1	83.8	79.4	73.5	72.8	67.6	64.0	63.2	64.7
	± 60	100	100	98.5	94.9	85.3	81.6	77.2	75.7	74.3	69.9	68.4	67.6
	± 120	100	100	100	98.5	91.2	87.5	84.6	81.6	80.9	75.7	72.1	73.5
	± 180	100	100	100	99.3	93.4	89.7	87.5	84.6	84.6	79.4	77.2	77.2
Technisch falsch	± 5	100	100	98.0	67.3	58.2	53.1	43.9	41.8	39.8	37.8	41.8	41.8
	± 10	100	100	98.0	88.8	67.3	60.2	53.1	48.0	41.8	41.8	44.9	44.9
	± 30	100	100	98.0	91.8	78.6	73.5	67.3	67.3	62.2	57.1	58.2	59.2
	± 60	100	100	98.0	92.9	80.6	76.5	71.4	70.4	68.4	62.2	62.2	62.2
	± 120	100	100	100	98.0	88.8	84.7	80.6	77.6	76.5	69.4	67.3	68.4
	± 180	100	100	100	99.0	91.8	87.8	84.7	81.6	81.6	74.5	73.5	73.5
Technisch richtig	± 5	100	100	94.7	71.1	68.4	65.8	65.8	71.1	71.1	71.1	68.4	71.1
	± 10	100	100	100	94.7	78.9	73.7	68.4	76.3	71.1	73.7	73.7	71.1
	± 30	100	100	100	100	97.4	94.7	89.5	86.8	81.6	81.6	76.3	78.9
	± 60	100	100	100	100	97.4	94.7	92.1	89.5	89.5	89.5	84.2	81.6
	± 120	100	100	100	100	97.4	94.7	94.7	92.1	92.1	92.1	84.2	86.8
	± 180	100	100	100	100	97.4	94.7	94.7	92.1	92.1	92.1	86.8	86.8

Tabelle 9.14: Sensitivitäten (in %) für verschiedene Fensterbreiten und Intervalle um einen als alarmrelevant annotierten Schwellwertalarm

sterbreiten werden bei noch relativ vielen Patienten vom neuen System mehr Alarme erzeugt (siehe Tabelle 9.13). Ab Fensterbreite 71 erhält man sowohl insgesamt eine lohnenswerte Reduzierung der Alarmhäufigkeit als auch zufriedenstellende Ergebnisse für die meisten Patienten.

Bei der Sauerstoffsättigung wird die Länge des für die Sensitivitätsbestimmung benötigten Intervalls um einen alarmrelevant annotierten Alarm auf ± 60 Sekunden festgelegt. Tabelle 9.14 enthält zusätzlich zum Vergleich auch Sensitivitäten für andere Intervalllängen. Der Alarm-Algorithmus mit Fensterbreite 71 erreicht in Bezug auf die Erkennung nicht technisch bedingter alarmrelevanter Situationen eine Sensitivität von 92.1%. Hier werden drei von insgesamt 38 alarmrelevanten Situationen nicht erkannt. Eine Situation weniger erkennen die Algorithmen mit den Fensterbreiten 81 bis 101. Von den 98 alarmrelevanten Zuständen aufgrund eines technischen Problems werden bei Verwendung von Fensterbreite 71 insgesamt 28 nicht erkannt.

Wird bei den signalextraktionsbasierten Alarm-Algorithmen auch eine Alarmbestätigungsfrist von zehn Sekunden angesetzt, erhält man zwar eine größere Verminderung der Alarmhäufigkeit, aber gleichzeitig auch deutlich schlechtere Sensitivitäten. Dassel-

	altes	Signalextraktion mit Fensterbreite (Sek.)			
	System	71	81	91	101
falsch positiv	223	178 (-20%)	157 (-30%)	152 (-32%)	151 (-32%)
PPV	39%	29%	32%	30%	29%

Tabelle 9.15: Anzahl Alarme außerhalb eines alarmrelevanten Zustandes (falsch positiv) und Positiver Prädiktiver Wert (PPV)

	Altes	Fensterbreite			
	System	71	81	91	101
Spezifität	87.7	90.2	91.3	91.6	91.7

Tabelle 9.16: Spezifitäten (in %) für SpO₂-Alarme

be ist der Fall, wenn die Signalschätzung für die Fenstermitte für die Alarmgebung verwendet wird. Alarm-Algorithmen auf Basis von Signalextraktion mittels adaptiver Fensterbreitenwahl (gewählte minimale Fensterbreite 60 bzw. 90, maximale Fensterbreite 600 und Entscheidung über Veränderung der Fenstergröße anhand der letzten 30 Werte im Fenster) erzielen ähnliche Ergebnisse wie die Algorithmen mit den festen Fensterbreiten 61 bzw. 91. Da keine Verbesserung erzielt wird, werden diese Ansätze hier nicht weiter verfolgt.

Die Anzahl falsch positiver Alarme und die Spezifitäten werden im Folgenden nur für die Fensterbreite 71, 81, 91 und 101 betrachtet, da nur für diese sowohl hinsichtlich Alarmreduktion als auch Sensitivität zufriedenstellende Ergebnisse erzielt werden. Tabelle 9.15 zeigt, dass für die ausgewählten Fensterbreiten zwischen 20% und 32% der falsch positiven Alarme vermieden werden. Wie bei den arteriellen Blutdrücken sind die Positiven Prädiktiven Werte der neuen Alarm-Algorithmen niedriger als der PPV des einfachen Schwellwertalarmsystems. Die Ursachen hierfür sind dieselben wie zuvor. Für die Bestimmung der Spezifität auf Basis der Alarmanzahl wird angenommen, dass schlimmstenfalls ein potentieller falscher Alarm pro zehn Minuten Beobachtungsdauer auftritt. Begründet wird diese Wahl wieder mit den im Referenzdatensatz vorkommenden SpO₂-Schwellwertalarmen pro Stunde für die einzelnen Patienten. Dort treten bei allen Fällen höchstens sechs einfache Schwellwertalarme pro Stunde auf, was einem Alarm alle zehn Minuten entspricht. Tabelle 9.16 zeigt die so bestimmten Werte für die Spezifitäten.

Die Analysen zeigen, dass sich trotz der zu Beginn erwähnten Argumente gegen signal-extraktionsbasierte Schwellwertalarme für die Sauerstoffsättigung bei genügend großer

Fensterbreite eine deutliche Reduzierung der Alarmhäufigkeit ohne großen Sensitivitätsverlust erreichen lässt. Allerdings sind die Sensitivitätswerte niedriger als für den Nicht-Unterlegenheitsnachweis gefordert. Dennoch soll ein neuer Alarm-Algorithmus auch auf den verbleibenden Daten des Referenzdatensatzes getestet werden. Da der Alarm-Algorithmus mit Fensterbreite 71 nur eine (technisch richtige) alarmrelevante Situation mehr erkennt als die Signalextraktionsalarme mit den Fensterbreiten 81-101, letztere aber wesentlich mehr überflüssige Alarme vermeiden, wird als Kompromiss der neue Alarm-Algorithmus mit Fensterbreite 91 für die Validierung ausgewählt. Für diesen finden sich in Tabelle C.8 die Ergebnisse der Vergleichsmethode von Charbonnier und Gentil (2007). Für vier der 114 unterdrückten alten Alarme wurde ein alarmrelevanter Zustand (bei technisch richtiger Datenaufzeichnung) annotiert. Dieses Ergebnis bestätigt das Resultat für die Sensitivität des Alarm-Algorithmus. 62 Alarme wurden zusätzlich ausgelöst. Teilweise wieder in Situationen, in denen sowohl das extrahierte Signal als auch die Originalzeitreihe unterhalb der unteren Alarmgrenze liegen, letztere aber für kürzere Zeit als die eingestellte Alarmbestätigungsfrist. Viele scheinbar zusätzliche neue Alarme treten aber auch in Situationen auf, in denen mehrere überwachte Vitalparameter durch eine Manipulation am Monitorsystem Werte außerhalb der Alarmgrenzen annehmen. Für das Infinity Monitoring System[®] wird in diesen Fällen möglicherweise der Alarm für einen anderen Vitalparameter als für die Sauerstoffsättigung registriert, da bei der Datenaufzeichnung keine „gleichzeitigen“ Alarme erfasst werden (siehe Abschnitt 8.1).

9.2 Validierung neuer Alarm-Algorithmen

Im Folgenden werden die ausgewählten Alarm-Algorithmen auf Basis der verbleibenden 45 Fälle des Referenzdatensatzes getestet. In Tabelle C.9 sind für die betreffenden Fälle jeweils für die arteriellen Blutdrücke, die Herzfrequenz und die Sauerstoffsättigung die Anzahl einfacher Schwellwertalarme insgesamt und pro Stunde sowie die Anzahl alarmrelevanter Zustände angegeben. Die Gesamt-Beobachtungsdauer beträgt knapp 647 Stunden.

Die Validierung erfolgt für die drei genannten Vitalparameter getrennt. Es werden somit nicht zwei komplette Alarmsysteme, sondern jeweils nur einzelne Alarmsystem-Komponenten miteinander verglichen. Dies ermöglicht eine differenzierte Untersuchung der Stärken und Schwächen der neuen Verfahren. Eine getrennte Betrachtung ist außerdem im Hinblick auf die gewählten Sensitivitäts- bzw. Spezifitätsdefinitionen sinnvoll,

da hierfür eine Vitalparameter-spezifische Aufteilung des Beobachtungszeitraums in alarmrelevante und nicht alarmrelevante Zeiträume vorgenommen wird. Eine alarmrelevante Situation für den arteriellen Blutdruck würde z.B. auch nur durch einen Alarm für ART.S oder ART.M erkannt.

Um zu entscheiden, ob die neu entwickelten Alarm-Algorithmen besser abschneiden als das einfache Schwellwertalarmsystem, wird die in Abschnitt 6.2.3 beschriebene Vorgehensweise gewählt. Beim Sensitivitätsvergleich wird die Sensitivität in Bezug auf die Erkennung alarmrelevanter Situationen bei technisch richtiger Datenaufzeichnung verwendet. Die Ergebnisse der Signifikanztests werden durch weitere Analysen hinsichtlich zusätzlicher Beurteilungskriterien ergänzt.

9.2.1 Arterielle Blutdrücke

Mit dem in Abschnitt 9.1.1 ausgewählten Alarm-Algorithmus werden insgesamt 816 Alarme für den systolischen und mittleren arteriellen Blutdruck generiert. Im Vergleich zum einfachen Schwellwertalarmsystem, bei dem 1477 Alarme auftreten, entspricht dies einer Reduzierung der Alarmhäufigkeit um 45%. Dies stimmt bis auf einen Prozentpunkt mit dem vorher erzielten Ergebnis für die Teilstichprobe des Referenzdatensatzes überein (siehe Abbildung C.4). Für 38 Fälle werden vom neuen Algorithmus weniger Alarme erzeugt, bei vier Fällen bleibt die Alarm-Anzahl unverändert und bei drei Fällen wird jeweils ein Alarm mehr generiert. Maximal werden bei den betrachteten Patienten neun Alarme pro Stunde weniger ausgelöst, im Mittel ein Alarm pro Stunde weniger.

In Bezug auf die Erkennung alarmrelevanter Zustände bei technisch richtiger Datenaufzeichnung wird eine Sensitivität von 90.7% erreicht. Von 182 alarmrelevanten Situationen werden 17 vom neuen System nicht rechtzeitig erkannt. Tabelle C.10 zeigt die Sensitivitäten getrennt für die einzelnen Patienten. Bei 16 Fällen tritt kein alarmrelevanter Zustand für die invasiven Blutdrücke auf, so dass für diese kein Sensitivitätswert bestimmt werden kann. Für 19 Fälle werden alle vorliegenden alarmrelevanten Zustände erkannt. Bei zehn Patienten erfolgt bei mindestens einer alarmrelevanten Situation kein signalextraktionsbasierter Alarm. Eine genauere Analyse der nicht erkannten alarmrelevanten Zustände ergibt, dass in zehn Situationen neue Alarme zwar mehr als 60 Sekunden, aber weniger als fünf Minuten Abstand zum alarmrelevant annotierten einfachen Schwellwertalarm haben und somit zumindest zeitnah erfolgen. In einer Situation ist das extrahierte Signal für ART.S zwar außerhalb der Alarmgrenzen, ein Alarm wurde aber nicht gezählt, da zum Zeitpunkt der Grenzüberschreitung noch

ein Alarm für ART.M aktiv war (siehe Regel zur Alarmauslösung in Abschnitt 8.1). Die übrigen sechs nicht erkannten alarmrelevant bewerteten Situationen sind in Abbildung C.6 dargestellt. In Abbildung C.6(a) und (b) kehrt die Originalzeitreihe – möglicherweise aufgrund einer erfolgten Intervention als Reaktion auf den Alarm – schnell wieder zu ihrem Ausgangsniveau zurück, so dass dem extrahierten Signal keine Gelegenheit gegeben wird, die alarmrelevante Situation zu erkennen. In den alarmrelevanten Situationen in Abbildung C.6(c) und (d) verletzt die Originalzeitreihe die Alarmgrenze nur minimal und das extrahierte Signal verläuft genau auf der Alarmgrenze. In C.6(d) wird zudem die Alarmgrenze genau dann heruntergestellt, als das extrahierte Signal die vorher geltende Alarmgrenze unterschreitet. In Abbildung C.6(e) werden die kurzfristigen Schwankungen der Originalzeitreihe vom extrahierten Signal nicht nachgezeichnet. Die alarmrelevante Situation in Abbildung C.6(f) kann vom neuen System nicht erkannt werden, da die Aufzeichnung kurz nach Auftreten der Situation endet.

Für sämtliche der nicht erkannten alarmrelevant bewerteten Situationen lässt sich nach Aussage eines an der Studie beteiligten Mediziners auf Basis der Zeitreihen der Vitalparameter die Aussage treffen, dass das Ausbleiben eines Alarms keine bedenklichen Konsequenzen für den Patienten gehabt hätte. Diese Feststellung ändert jedoch nichts an dem oben angegebenen Sensitivitätswert, da die Auswertung dieser Studie an die vorher gewählte Definition eines alarmrelevanten Zustandes und die Methode der Sensitivitätsbestimmung gebunden ist.

Ein Alarm-Vergleich mittels der Methode von Charbonnier und Gentil (2007) ergibt, dass 16 Alarme mit Annotation „alarmrelevant und technisch richtig“ unterdrückt werden (siehe Tabelle C.13). Dies stimmt bis auf einen Alarm mit der bei der Sensitivitätsbestimmung erhaltenen Anzahl nicht erkannter alarmrelevanter Zustände überein.

Die Sensitivität in Bezug auf die Erkennung alarmrelevanter Zustände aufgrund eines technischen Problems beträgt 82.4%. Dieser Wert ist im Vergleich zu dem in Abschnitt 9.1.1 für den anderen Teil des Referenzdatensatzes erhaltenen Wert deutlich besser. Insgesamt resultiert so eine Sensitivität von 87.9%.

Das einfache Schwellwertalarmsystem löst 1130 falsch positive Alarme aus. Der Positive Prädiktive Wert beträgt 23%. Der neue Alarm-Algorithmus erreicht mit 626 falsch positiven Alarmen eine Reduzierung der Alarmhäufigkeit um 45%. Auch hier erhält man einen Positiven Prädiktiven Wert von 23%. Die Anzahl der falsch positiven Alarme für die einzelnen Fälle findet sich in Tabelle C.10. Hier sind auch die Einzel-Spezifitäten und Gesamt-Spezifitäten für den alten und neuen Alarm-Algorithmus angegeben. Das einfache Schwellwertalarmsystem erreicht eine Spezifität von 87.2%. Für den neuen Alarm-Algorithmus beträgt die Spezifität 92.9%.

Sensitivitäts-Vergleich

$\hat{\text{Se}}(A_{alt})$:	100% (fest)
$\hat{\text{Se}}(A_{neu})$:	90.7% (95%-KI: [86.2%, 95.1%])
Testproblem:	$H_0^{\text{Se}} : \text{Se}(A_{neu}) \leq 0.95$ vs. $H_1^{\text{Se}} : \text{Se}(A_{neu}) > 0.95$
Teststatistik:	$X_{W''} = -1.906$
Testentscheidung:	$-1.906 < u_{1-0.05} = 1.64$ (p-Wert: 0.97) $\Rightarrow H_0^{\text{Se}}$ kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden.

Spezifitäts-Vergleich

$\hat{\text{Sp}}(A_{alt})$:	87.2% (95%-KI: [83.0%, 91.5%])
$\hat{\text{Sp}}(A_{neu})$:	92.9% (95%-KI: [91.0%, 94.9%])
Testproblem:	$H_0^{\text{Sp}} : \text{Sp}(A_{alt}) - \text{Sp}(A_{neu}) \geq 0$ vs. $H_1^{\text{Sp}} : \text{Sp}(A_{alt}) - \text{Sp}(A_{neu}) < 0$
Teststatistik:	$X_O = -3.507$
Testentscheidung:	$-3.507 < u_{0.05} = -1.64$ (p-Wert: 0.0002) $\Rightarrow H_0^{\text{Sp}}$ wird zum Niveau $\alpha = 0.05$ abgelehnt .

Simultaner Vergleich

IUT-Testproblem:	$H_0 : H_0^{\text{Se}} \cup H_0^{\text{Sp}}$ vs. $H_1 : H_1^{\text{Se}} \cap H_1^{\text{Sp}}$
IUT-Testentscheidung:	H_0 kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden, da H_0^{Se} nicht abgelehnt werden kann.

Tabelle 9.17: Validierung des Alarm-Algorithmus für die arteriellen Blutdrücke

Die für die Validierung interessierenden Test-Ergebnisse sind in Tabelle 9.17 zusammengefasst. Für die Sensitivität bezüglich der Erkennung alarmrelevanter Zustände ohne technische Ursache kann die Nicht-Unterlegenheit im Vergleich zum alten System für die gewählte untere Schranke von 95%, wie erwartet, zum Signifikanzniveau von 5% nicht gezeigt werden. Die Überlegenheit in Bezug auf die Spezifität kann jedoch nachgewiesen werden. Da im insgesamt zu überprüfenden simultanen Intersection-Union-Testproblem jedoch Sensitivitäts- und Spezifitäts-Nullhypothese nicht gleichzeitig abgelehnt werden können, kann keine Überlegenheit des neuen Alarm-Algorithmus demonstriert werden.

9.2.2 Herzfrequenz

Der in Abschnitt 9.1.2 für die Herzfrequenz ausgewählte neue Alarm-Algorithmus generiert für die vorliegenden Daten insgesamt 180 Alarme. Im Vergleich zu den 246 Alarmen des einfachen Schwellwertsystems sind dies 27% weniger. Für fünf Patienten werden vom neuen Verfahren mehr Alarme ausgelöst, bei einem Patienten ist die

Differenz der Alarmanzahlen mit elf Alarmen relativ groß. Bei 26 Fällen bleibt die Alarmhäufigkeit unverändert, was in 21 dieser Fälle dadurch begründet ist, dass kein einfacher Schwellwertalarm für die Herzfrequenz vorliegt. Für 14 Fälle generiert der neue Alarm-Algorithmus weniger Alarme.

Die Anzahl der alarmrelevanten Zustände ist mit 21 sehr gering. Sie teilt sich auf nur neun der 45 betrachteten Patienten auf. Alle alarmrelevanten Zustände beziehen sich auf Situationen mit technisch richtiger Datenaufzeichnung. Der neue Alarm-Algorithmus erkennt nur elf dieser Situationen, so dass die Sensitivität in diesem Fall nur 52.4% beträgt (siehe auch Tabelle C.11 für eine Übersicht über die Sensitivitäten der einzelnen Patienten). Für eine der nicht erkannten Situationen erfolgt ein neuer Alarm zwar mehr als die noch zulässigen 30 Sekunden zu spät, der Abstand ist aber mit 54 Sekunden noch akzeptabel. Eine genauere Analyse der übrigen neun nicht erkannten alarmrelevanten Zustände zeigt, dass bei den meisten Situationen das eigentliche Niveau der HR-Zeitreihe zwar deutlich innerhalb der Grenzen liegt, die Variabilität aber sehr groß ist, so dass durch die Schwankungen die Alarmgrenze kurzzeitig und minimal verletzt wird. Die Schwankungen werden vom extrahierten Signal abgeschwächt, so dass kein Alarm erfolgt (siehe Abbildung C.7). Bei mehreren der betreffenden alarmrelevanten Situationen werden vom Monitor zeitgleich Arrhythmie-Ereignisse (z.B. Tachykardie, Bradykardie) identifiziert. Arrhythmie-Alarme anstelle von Schwellwertalarmen wären in diesen Fällen aussagekräftiger.

Von den mit der Methode von Charbonnier und Gentil (2007) identifizierten unterdrückten Alarmen des einfachen Schwellwertalarmsystems sind neun mit „alarmrelevant und technisch richtig“ annotiert (siehe Tabelle C.14). Dies bestätigt bis auf einen Alarm die oben angegebene Anzahl nicht erkannter alarmrelevanter Situationen.

Die Anzahl der falsch positiven Alarme für das einfache Schwellwertalarmsystem beträgt 222 (PPV = 10%). Das neue Alarmgebungs-Verfahren erzeugt 168 – d.h. 24% weniger – falsch positive Alarme (PPV = 7%). Die Spezifitäten für den alten und den neuen Alarm-Algorithmus sind mit Werten von 93.8% bzw. 95.3% zum einen recht hoch, was unter anderem dadurch zu erklären ist, dass bei vielen Fällen überhaupt keine HR-Schwellwertalarme auftreten. Zum anderen unterscheiden sich die Spezifitäten auch nur geringfügig (Tabelle C.11 enthält ergänzend die Spezifitäten für die einzelnen Fälle).

In Tabelle 9.18 sind die Ergebnisse der Signifikanz-Tests für den Sensitivitäts- und Spezifitätsvergleich zusammengefasst. Da zum 5%-Niveau weder die Nicht-Unterlegenheit bezüglich der Sensitivität noch die Überlegenheit bezüglich der Spezifität gezeigt werden kann, wird auch insgesamt die Überlegenheit des neuen Alarm-Algorithmus für die

Sensitivitäts-Vergleich

$\hat{\text{Se}}(A_{alt})$:	100% (fest)
$\hat{\text{Se}}(A_{neu})$:	52.4% (95%-KI: [30.5%, 74.2%])
Testproblem:	$H_0^{\text{Se}} : \text{Se}(A_{neu}) \leq 0.95$ vs. $H_1^{\text{Se}} : \text{Se}(A_{neu}) > 0.95$
Teststatistik:	$X_{W''} = -3.823$
Testentscheidung:	$-3.823 < u_{1-0.05} = 1.64$ (p-Wert: 0.99) $\Rightarrow H_0^{\text{Se}}$ kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden.

Spezifitäts-Vergleich

$\hat{\text{Sp}}(A_{alt})$:	93.8% (95%-KI: [90.0%, 97.7%])
$\hat{\text{Sp}}(A_{neu})$:	95.3% (95%-KI: [92.2%, 98.4%])
Testproblem:	$H_0^{\text{Sp}} : \text{Sp}(A_{alt}) - \text{Sp}(A_{neu}) \geq 0$ vs. $H_1^{\text{Sp}} : \text{Sp}(A_{alt}) - \text{Sp}(A_{neu}) < 0$
Teststatistik:	$X_O = -1.092$
Testentscheidung:	$-1.092 > u_{0.05} = -1.64$ (p-Wert: 0.14) $\Rightarrow H_0^{\text{Sp}}$ kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden.

Simultaner Vergleich

IUT-Testproblem:	$H_0 : H_0^{\text{Se}} \cup H_0^{\text{Sp}}$ vs. $H_1 : H_1^{\text{Se}} \cap H_1^{\text{Sp}}$
IUT-Testentscheidung:	H_0 kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden, da H_0^{Se} und H_0^{Sp} nicht abgelehnt werden können.

Tabelle 9.18: Validierung des Alarm-Algorithmus für die Herzfrequenz

Herzfrequenz nicht demonstriert. Es ist dabei zu berücksichtigen, dass die Aussagekraft des Ergebnisses für den Sensitivitätsvergleich durch die vergleichsweise geringe Anzahl alarmrelevanter Zustände geschmälert wird.

9.2.3 Sauerstoffsättigung

Der in Abschnitt 9.1.3 ausgewählte signalextraktionsbasierte Alarm-Algorithmus für die Sauerstoffsättigung löst für die vorliegenden Daten insgesamt 303 Alarme aus. So wird gegenüber dem einfachen Schwellwertalarmsystem – welches 419 Alarme generiert – eine Reduzierung der Alarmhäufigkeit um 28% bewirkt. Im Vergleich zu dem Ergebnis für die in Abschnitt 9.1.3 betrachteten Patienten ist die Alarmreduzierung hier um 12 Prozentpunkte geringer. Bei neun Fällen werden mehr Alarme generiert, wobei die Anzahl zusätzlicher Alarme klein ist. Für 14 Fälle unterscheiden sich die Alarmhäufigkeiten für den alten und neuen Alarm-Algorithmus nicht. In 22 Fällen werden vom neuen Verfahren weniger Alarme ausgelöst.

Das neue alarmgebende Verfahren erkennt 55 der insgesamt 61 alarmrelevanten

Sensitivitäts-Vergleich

$\hat{Se}(A_{alt})$:	100% (fest)
$\hat{Se}(A_{neu})$:	90.2% (95%-KI: [79.7%, 100%])
Testproblem:	$H_0^{Se} : Se(A_{neu}) \leq 0.95$ vs. $H_1^{Se} : Se(A_{neu}) > 0.95$
Teststatistik:	$X_{W''} = -0.906$
Testentscheidung:	$-0.906 < u_{1-0.05} = 1.64$ (p-Wert: 0.82) $\Rightarrow H_0^{Se}$ kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden.

Spezifitäts-Vergleich

$\hat{Sp}(A_{alt})$:	92.0% (95%-KI: [88.9%, 95.0%])
$\hat{Sp}(A_{neu})$:	93.6% (95%-KI: [91.4%, 95.7%])
Testproblem:	$H_0^{Sp} : Sp(A_{alt}) - Sp(A_{neu}) \geq 0$ vs. $H_1^{Sp} : Sp(A_{alt}) - Sp(A_{neu}) < 0$
Teststatistik:	$X_O = -2.057$
Testentscheidung:	$-2.057 < u_{0.05} = -1.64$ (p-Wert: 0.02) $\Rightarrow H_0^{Sp}$ wird zum Niveau $\alpha = 0.05$ abgelehnt .

Simultaner Vergleich

IUT-Testproblem:	$H_0 : H_0^{Se} \cup H_0^{Sp}$ vs. $H_1 : H_1^{Se} \cap H_1^{Sp}$
IUT-Testentscheidung:	H_0 kann zum Niveau $\alpha = 0.05$ nicht abgelehnt werden, da H_0^{Se} nicht abgelehnt werden kann.

Tabelle 9.19: Validierung des Alarm-Algorithmus für die Sauerstoffsättigung

Zustände, die sich nicht auf ein technisches Problem beziehen, so dass eine Sensitivität von 90.2% resultiert. In Tabelle C.12 finden sich die Sensitivitäten für die einzelnen Patienten. Bei den nicht erkannten alarmrelevanten Zuständen handelt es sich ausschließlich um Situationen, in denen die Alarmgrenze nur für einen kurzen Zeitraum und teilweise auch nur geringfügig unterschritten wird (siehe auch Abbildung C.8) und die auf Basis der Zeitreihen aus medizinischer Sicht für den Patienten keine schweren gesundheitlichen Folgen befürchten lassen. Auf die Sensitivität wirkt sich dies aber nicht aus, da für die Validierung die in den Annotationen auf Basis des Videos identifizierten alarmrelevanten Zustände maßgeblich sind.

Die Anzahl unterdrückter Alarme, ermittelt nach Charbonnier und Gentil (2007), mit Annotation „alarmrelevant und technisch richtig“ stimmt mit der bei der Sensitivitätsbestimmung erhaltenen Anzahl nicht erkannter alarmrelevanter Zustände überein (siehe Tabelle C.15).

Die Sensitivität in Bezug auf die Erkennung alarmrelevanter Zustände technischer Ursache beträgt 90.5%. Nur sieben der 74 alarmrelevanten Situationen werden nicht erkannt. Somit ergibt sich eine Gesamt-Sensitivität von 90.4%.

Die Anzahl falsch positiver Alarme für das einfache Schwellwertalarmsystem beträgt

270 (PPV = 36%). Der neue Alarm-Algorithmus erzeugt 216 falsch positive Alarme (PPV = 29%) und leistet somit eine Reduzierung um 20%. Als Spezifität für das alte Alarmsystem erhält man 92.0%. Der neue Alarm-Algorithmus weist eine Spezifität von 93.6% auf.

Die Zusammenfassung der Testentscheidungen für den Sensitivitäts- und Spezifitätsvergleich findet sich in Tabelle 9.19. Die Nicht-Unterlegenheit des neuen Alarm-Algorithmus in Bezug auf die Sensitivität kann zum 5%-Niveau nicht nachgewiesen werden. Obwohl der signalextraktionsbasierte Alarm-Algorithmus hinsichtlich der Spezifität signifikant besser abschneidet als das einfache Schwellwertalarmsystem, kann somit insgesamt keine Überlegenheit des neuen Verfahrens gezeigt werden.

9.3 Diskussion der Ergebnisse

Für die hier untersuchten Alarm-Algorithmen kann anhand der vor der Datenerhebung ausgewählten Beurteilungskriterien Folgendes gezeigt werden: Die neuen Alarm-Algorithmen für die arteriellen Blutdrücke und die Sauerstoffsättigung sind zwar hinsichtlich der Spezifität deutlich besser als das einfache Schwellwertalarmsystem, die Nicht-Unterlegenheit bezüglich der Sensitivität ist aber für den vorher gewählten noch akzeptablen Sensitivitätsunterschied von 5% nicht nachweisbar. Der Alarm-Algorithmus für die Herzfrequenz erreicht für die Validierungsdaten eine sehr niedrige Sensitivität und dies ohne hinsichtlich der Spezifität signifikant besser abzuschneiden als das alte Alarmsystem. Allerdings ist bei der Herzfrequenz zum einen das Potential für eine Alarmreduzierung eher gering. Zum anderen stehen nur wenige alarmrelevante Zustände als Entscheidungsgrundlage zur Verfügung, die zudem teilweise die Besonderheit aufweisen, dass sie adäquater durch einen Arrhythmie-Alarm als durch einen einfachen oder signalextraktionsbasierten Schwellwertalarm angezeigt werden würden. Wie bereits vorher angesprochen, ist das Abschneiden hinsichtlich der Sensitivität teilweise auf die gewählte Definition eines alarmrelevanten Zustandes zurückzuführen. Diese besagt, dass auf einen alarmrelevanten Zustand (der sich nicht auf ein technisches Problem bezieht) eine diagnostische oder therapeutische Entscheidung erfolgen muss. Wenn im Video eine entsprechende Handlung zu beobachten ist, können somit auch Situationen, in denen die Zeitreihenwerte allein nicht auf eine kritische Veränderung des Patientenzustandes schließen lassen, als alarmrelevant bewertet werden. Für Teile der vorliegenden Daten und mit Hilfe von Klassifikationsverfahren zeigt Brinschwitz (2007) im Rahmen seiner Diplomarbeit, dass allein auf Basis von Alarm-Charakteristika wie

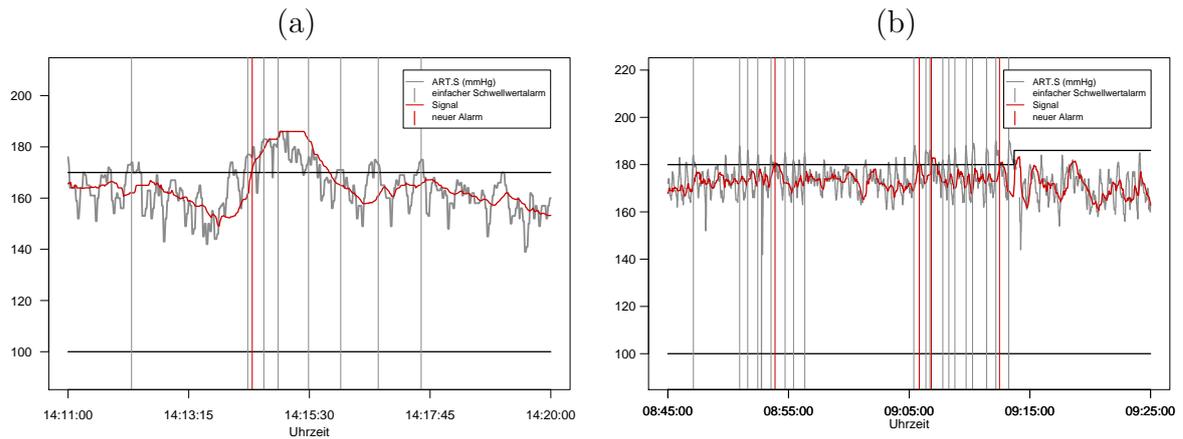


Abbildung 9.10: Beispiele für vorteilhaftes Verhalten signalextraktionsbasierter Alarm-Algorithmen: (a) Auslösung eines langen Alarms statt vieler kurzer Alarme (b) Weniger Alarme bei eng eingestellten Alarmgrenzen

z.B. Dauer und Ausmaß der Grenzverletzungen keine zuverlässigen Aussagen über die Alarmrelevanz der zugrunde liegenden Situation gemacht werden können. In einigen Fällen wird der natürliche Verlauf der Zeitreihe möglicherweise durch eine Intervention beeinflusst, so dass in der Zeitreihe nur kurze und abrupte Veränderungen sichtbar sind. In solchen Fällen hat ein signalextraktionsbasierter Alarm-Algorithmus, der auf dem Prinzip des Filterns gerade solcher Veränderungen beruht, bei der Offline-Anwendung keine Chance die alarmrelevante Situation zu erkennen.

Unter Berücksichtigung dieser Punkte ist eine Sensitivität von 90%, die bei den Alarm-Algorithmen für die arteriellen Blutdrücke und die Sauerstoffsättigung erreicht wird, möglicherweise aus medizinischer Sicht noch hoch genug, wenn die damit einhergehende Reduzierung der Alarmhäufigkeit außergewöhnlich gut ist. Auch die Sensitivität bezüglich der Erkennung dauerhafter technischer Probleme, welche bei der Validierung nicht berücksichtigt wurde, liegt für beide Alarm-Algorithmen über 80%.

Bei der Sauerstoffsättigung rechtfertigt eine Alarmreduktion von 28% im Hinblick auf alle Alarme und von 20% bezüglich der falsch positiven Alarme den Sensitivitätsverlust unter Umständen nicht. Anders sieht dies beim neuen Alarm-Algorithmus für die arteriellen Blutdrücke aus, mit dem 45% der Alarme (insgesamt und falsch positiv) vermieden werden. Besonders deutlich werden die Vorteile des neuen Verfahrens bei Betrachtung von Abbildung 9.10. Der signalextraktionsbasierte Alarm-Algorithmus löst z.B. eher einen langen Alarm statt viele kurze Alarme aus (Abbildung 9.10(a)). Schwankt die Originalzeitreihe längere Zeit um die (möglicherweise zu eng eingestellte) Alarmgrenze, werden vom neuen Verfahren außerdem deutlich weniger Alarme als vom einfachen Schwellwertalarmsystem generiert (Abbildung 9.10(b)).

Einfache Schwellwertalarme in Situationen wie in Abbildung 9.10 tragen auch häufig

die Annotation „hinweisend“, d.h. die Information über die Verletzung der Alarmgrenze ist nützlich, es ist aber keine unmittelbare Handlung erforderlich. In Abschnitt 6.3 wurde bereits darauf hingewiesen, dass es neben der Erkennung alarmrelevanter Zustände auch von Interesse ist, ob der neue Alarm-Algorithmus in Situationen, in denen „hinweisende“ alte Alarmer auftreten, ebenfalls alarmiert – nur möglicherweise mit weniger Alarmen. Hierzu können z.B. die in Abschnitt 7.4 identifizierten Gruppen einfacher Schwellwertalarmer für die arteriellen Blutdrücke mit gleicher Alarmursache (siehe Tabelle 7.4) betrachtet werden. Für jede Gruppe, die mindestens einen hinweisenden Alarm enthält und jeden vereinzelt auftretenden hinweisenden Alarm, kann dann geprüft werden, ob in einem Zeitraum von ± 60 Sekunden um die Alarm-Gruppe oder den Einzel-Alarm auch ein neuer Alarm auftritt. Von den 641 untersuchten Situationen mit hinweisenden Alarmen werden 500 (78%) durch den neuen Alarm-Algorithmus ebenfalls alarmiert. Bei den nicht alarmierten Situationen handelt es sich überwiegend um Fälle, in denen nur ein hinweisender alter Alarm auftritt. Für die alarmierten Situationen werden in 3% der Fälle mehr, in 36% der Fälle gleich viele und in 61% der Fälle weniger neue Alarmer gegeben. Die Anzahl der weniger ausgelösten Alarmer in einer Situation beträgt im Mittel zwei und maximal 15 Alarmer. Der neue Alarm-Algorithmus schneidet somit auch in dieser Hinsicht gut ab.

Neben der Datenerhebung für den Referenzdatensatz wurden am Studienort Regensburg für eine große Anzahl Patienten auch Daten von Infinity Monitoren[®] gesammelt, welche nicht annotiert wurden. Für diese Daten kann zumindest in Bezug auf die Alarmreduktion überprüft werden, ob der neue Alarm-Algorithmus für die arteriellen Blutdrücke ähnlich abschneidet wie auf den Daten des Referenzdatensatzes. Von den nicht annotierten Daten werden dabei nur die Fälle betrachtet, bei denen der arterielle Blutdruck invasiv überwacht wurde. Insgesamt handelt es sich dabei um 180 Fälle mit einer Gesamt-Beobachtungsdauer von 3947 Stunden. Das einfache Schwellwertalarmsystem löst in diesem Zeitraum 6651 Alarmer für den systolischen und mittleren arteriellen Blutdruck aus. Bei der Offline-Anwendung des neuen Alarm-Algorithmus werden 3795, d.h. 43% weniger Alarmer generiert. Die erreichte Alarmreduktion ist somit nur unwesentlich kleiner als die auf Basis des Referenzdatensatzes erzielte Verminderung der Alarmhäufigkeit.

Kapitel 10

Zusammenfassung und Ausblick

In dieser Arbeit wird die Eignung neuer Alarm-Algorithmen für das intensivmedizinische Online-Monitoring untersucht. Hierfür wurde eine klinische Studie zur Erstellung eines klinisch annotierten Referenzdatensatzes geplant und durchgeführt, sowie eine statistische Validierungsmethodik entwickelt. Die neuen Verfahren wurden offline auf Basis des Referenzdatensatzes getestet.

Die neuen Alarm-Algorithmen basieren auf einer Vorverarbeitung der Vitalparameter-Zeitreihen mittels robuster Online-Signalextraktion. Ein Vergleich der extrahierten Signale mit Alarmgrenzen entscheidet über die Auslösung von Alarmen. Diese Vorgehensweise wird dem derzeit von Monitoring-Geräten verwendeten einfachen Schwellwert-alarmsystem gegenübergestellt, bei dem die Originalzeitreihen mit den Alarmgrenzen verglichen werden und welches viele Fehlalarme produziert. Das Datenerhebungskonzept der klinischen Studie stellt sicher, dass alle benötigten Informationen für einen Offline-Vergleich alter und neuer Alarm-Algorithmen zur Verfügung stehen. So umfasst der Referenzdatensatz nicht nur die sekundlich aufgezeichneten Werte aller gemessenen Vitalparameter, sondern auch die Alarme und Alarmeinstellungen des derzeit verwendeten Alarmsystems. Ergänzt werden die Daten vom Monitorsystem durch klinische Annotationen. Hierbei handelt es sich um Bewertungen der aufgetretenen Alarmsituationen auf Basis von Videoaufzeichnungen. In den Annotationen ist unter anderem das Vorliegen eines sogenannten alarmrelevanten Zustandes zum Zeitpunkt des Alarms festgehalten. Ein solcher ist in dieser Studie definiert als ein Zustand, auf den eine diagnostische oder therapeutische Entscheidung erfolgt oder der die Behebung eines dauerhaften technischen Problems erforderlich macht. Der vollständige Referenzdatensatz enthält Daten für 68 Fälle mit einer Gesamt-Beobachtungsdauer von 982 Stunden.

Als Haupt-Bewertungskriterien für die zu vergleichenden Alarmsysteme werden die Sensitivität bezüglich der Erkennung alarmrelevanter Zustände und die Spezifität, als Maß für die Fehlalarm-Anfälligkeit eines Alarmsystems, ausgewählt. Die Überlegenheit neuer Alarm-Algorithmen wird gezeigt, wenn zusätzlich zur Nicht-Unterlegenheit hinsichtlich der Sensitivität die Überlegenheit bezüglich der Spezifität nachgewiesen wird. Dabei wird die Nicht-Unterlegenheitsschranke für den Sensitivitätsunterschied auf 5% festgesetzt. Zur Bestimmung von Sensitivitäten ist die Bestimmung alarmrelevanter *Zeiträume* auf Basis der Annotationen für die *Alarmzeitpunkte* erforderlich. Für die Spezifitätsbestimmung wird ein Ansatz gewählt, bei dem die Anzahl der Fehlalarme berücksichtigt wird. Der Spezifitätsunterschied zwischen altem und neuem Alarmsystem kann sinnvoll für eine Testentscheidung darüber verwendet werden, welches System dem anderen überlegen ist. Die Betrachtung der Spezifitätswerte ist hingegen wenig aussagekräftig. Daher wird die Fehlalarm-Anfälligkeit auch immer über die tatsächliche Fehlalarm-Anzahl quantifiziert. Bei der Auswahl der Signifikanztests für die zu untersuchenden Testprobleme zum Vergleich von Sensitivitäten und Spezifitäten wird berücksichtigt, dass Abhängigkeiten zwischen den Beobachtungen bzw. Beobachtungspaaren eines Patienten vorliegen können. Der Gefahr antikonservativer Testentscheidungen wird so entgegengewirkt.

Für den Vergleich einfacher und signalextraktionsbasierter Schwellwertalarme wird zunächst eine Aufteilung des Referenzdatensatzes vorgenommen. Auf Basis einer Teilstichprobe wird untersucht, für welche Vitalparameter neue Alarme lohnenswert sind, welche Fensterbreite für die Signalextraktion jeweils geeignet ist und ob zusätzliche Alarmregeln erforderlich sind. Es zeigt sich, dass die Verwendung des reinen Online-Signals zur Alarmgebung nicht geeignet ist, da Über- und Unterschätzungen zu unerwünschten zusätzlichen Alarmen führen. Eine Beschränkung des Signals auf Werte zwischen Minimum und Maximum der Originalwerte im Fenster löst dieses Problem. Weiterhin können viele Alarme aufgrund minimaler Signalschwankungen um die Alarmgrenze vermieden werden, indem ein Alarm so lange aktiv bleibt, bis sich das extrahierte Signal wieder ausreichend weit innerhalb der Alarmgrenzen bewegt. Auf Basis der Teilstichprobe stellt sich für die arteriellen Blutdrücke und die Sauerstoffsättigung die Fensterbreite 91 Sekunden und für die Herzfrequenz die Fensterbreite 31 Sekunden als am geeignetsten für die signalextraktionsbasierte Alarmgebung heraus.

Bei einer Anwendung auf die übrigen Daten des Referenzdatensatzes sind die neuen Alarm-Algorithmen für die arteriellen Blutdrücke und die Sauerstoffsättigung dem einfachen Schwellwertalarmsystem zwar hinsichtlich der Spezifität deutlich überlegen, die Nicht-Unterlegenheit bezüglich der Sensitivität kann aber für den vor Beginn der Studie

gewählten noch akzeptablen Sensitivitätsunterschied von 5% nicht nachgewiesen werden. Für die neuen Herzfrequenz-Alarme wird weder die Nicht-Unterlegenheit bezüglich der Sensitivität noch die Überlegenheit bezüglich der Spezifität gezeigt. Hier ist allerdings zu berücksichtigen, dass aufgrund der eher geringen Alarmfrequenz für diesen Vitalparameter ohnehin kaum Alarme reduziert werden können und dass zudem die Aussagen bezüglich der Sensitivität auf nur sehr wenigen alarmrelevanten Zuständen basieren.

Die Alarm-Algorithmen für die arteriellen Blutdrücke und die Sauerstoffsättigung erreichen Sensitivitäten von knapp über 90%, während das einfache Schwellwertalarmsystem eine Sensitivität von 100% aufweist. Die Ursache hierfür liegt hauptsächlich darin, dass aufgrund der im Video beobachteten Handlungen teilweise auch Situationen als alarmrelevant identifiziert wurden, in denen die Zeitreihenwerte so unauffällig sind, dass die Veränderungen durch die robuste Signalextraktion gefiltert und somit nicht alarmiert werden. Zumindest für die Blutdruckalarme ist die (Fehl-)Alarmreduzierung jedoch mit 45% so deutlich, dass der Sensitivitätsverlust möglicherweise medizinisch doch noch akzeptabel ist. Diesbezüglich sei insbesondere darauf hingewiesen, dass laut Referenzdatensatz Blutdruck-Schwellwertalarme derzeit 57% der Schwellwertalarme und 40% aller Alarme ausmachen. Die erzielte Alarmreduktion würde sich somit auch insgesamt deutlich bemerkbar machen, wodurch die zu Beginn der Arbeit erwähnte Gefahr der Desensibilisierung gegenüber wichtigen Alarmen vermindert werden könnte.

Der im Rahmen der Studie erstellte annotierte Referenzdatensatz kann zur Überprüfung weiterer Verfahren zur Verbesserung der bestehenden Alarmsysteme verwendet werden. Dazu gehören beispielsweise Alarm-Algorithmen basierend auf multivariaten Signalextraktionsmethoden oder mittels Klassifikationsverfahren aus den Daten „gelernte“ Alarmregeln. Anhand des Referenzdatensatzes kann auch untersucht werden, ob Methoden zur Online-Überwachung der Variabilität für die Erkennung alarmrelevanter Zustände nützlich sind. Im Hinblick auf die Früherkennung solcher Zustände können auch Verfahren zur Trenderkennung – z.B. auf Basis der fensterweise berechneten Repeated Median Steigung – getestet werden. Zur Beurteilung dieser Methoden kann die in dieser Arbeit vorgestellte Validierungsmethodik herangezogen werden.

Für die Zukunft ist auch die Erhebung weiterer Daten auf Intensivstationen anderen Typs (chirurgisch, interdisziplinär, pädiatrisch,...) sinnvoll, um die Eignung neuer Methoden für ein breites Spektrum von Patienten zeigen zu können. Die in der Planung, Durchführung und Auswertung dieser Studie gemachten Erfahrungen liefern wichtige Erkenntnisse für solche und weitere zukünftige Studien, welche sicherlich nötig sind, um die Überführung neuer Methoden in die klinische Praxis voranzutreiben.

Literaturverzeichnis

- ABEL, U. (1993): *Die Bewertung diagnostischer Tests*, Hippokrates Verlag, Stuttgart.
- AGRESTI, A. (2002): *Categorical Data Analysis*, 2. Auflage, Wiley, New York.
- AHN, C. (1997): An Evaluation of Methods for the Estimation of Sensitivity and Specificity of Site-Specific Diagnostic Tests, *Biometrical Journal*, **39**, 793-807.
- BERGER, R.L., HSU, J.C. (1996): Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets, *Statistical Science*, **11**, 283-319.
- BIOT, L., CARRY, P.Y., PERDRIX, J.P., EBERHARD, A., BACONNIER P. (2000): Évaluation Clinique de la Pertinence des Alarmes en Réanimation, *Annales Françaises d'Anesthésie et de Réanimation*, **19**, 459-466.
- BRINSCHWITZ, H. (2007): *Statistische Analyse der klinischen Relevanz von intensivmedizinischen Schwellwertalarmen*, Diplomarbeit, Fakultät Statistik, Technische Universität Dortmund.
- CHAMBRIN, M.-C., RAVAUX, P., CALVELO-AROS, D., JABORSKA, A., CHOPIN, C., BONIFACE, B. (1999): Multicentric Study of Monitoring Alarms in the Adult Intensive Care Unit (ICU): A Descriptive Analysis, *Intensive Care Medicine*, **25**, 1360-1366.
- CHAMBRIN, M.-C., BIJAOU, E., CARRY, P.-Y., CHARBONNIER, S., DOJAT, M., GARBAY, C., MANGALABOYI, J., PERDRIX, J.-P., RAVAUX, P., SHARSHAR, S., SILVENT, A.-S., VILHELM, C. (2003): OSSCAR: A Collaborative Project for Intelligent Patient Monitoring in Intensive Care and Anesthesia, In: *Proceedings of the Workshop on Intelligent Data Analysis In Medicine and Pharmacology (IDAMAP) 2003 in Protaras, Zypern*.
- CHARBONNIER, S., BECQ, G., BIOT, L. (2004): On-line Segmentation Algorithms for Continuously Monitored Data in Intensive Care Units, *IEEE Transactions on Biomedical Engineering*, **51**, 484-492.

- CHARBONNIER, S., GENTIL, S. (2007): A Trend-based Alarm System to Improve Patient Monitoring in Intensive Care Patients, *Control Engineering Practice*, **15**, 1039-1050.
- CHEN, J.J., HSUEH, H.-M., LIU, J.-P. (2003): Simultaneous Non-inferiority Test of Sensitivity and Specificity for Two Diagnostic Procedures in the Presence of a Gold Standard, *Biometrical Journal*, **1**, 47-60.
- COCHRAN, W.G. (1977): *Sampling Techniques*, 3. Auflage, Wiley, New York.
- DAVIES, L.P., FRIED, R., GATHER, U. (2004): Robust Signal Extraction for On-line Monitoring Data, *Journal of Statistical Planning and Inference*, **122**, 65-78.
- DONNER, A., KLAR, N. (1993): Confidence Interval Construction for Effect Measures Arising from Cluster Randomization Trials, *Journal of Clinical Epidemiology*, **46**, 123-131.
- DRÄGER (2003): *Serie Infinity Delta Gebrauchsanweisung*, Dräger Medical Systems Inc., Danvers.
- DURKALSKI, V.L., PALESCH, Y.Y., LIPSITZ, S.R., RUST, P.F. (2003a): Analysis of Clustered Matched-Pair Data, *Statistics in Medicine*, **22**, 2417-2428.
- DURKALSKI, V.L., PALESCH, Y.Y., LIPSITZ, S.R., RUST, P.F. (2003b): Analysis of Clustered Matched-Pair Data for a Non-Inferiority Study Design, *Statistics in Medicine*, **22**, 279-290.
- ELIASZIW, M., DONNER, A. (1991): Application of the McNemar Test to Non-Independent Matched Pair Data, *Statistics in Medicine*, **10**, 1981-1991.
- FLEISS, J.L. (1981): *Statistical Methods for Rates and Proportions*, 2. Auflage, Wiley, New York.
- FRIED, R. (2004): Robust Filtering of Time Series with Trends, *Journal of Nonparametric Statistics*, **16**, Special Issue on the International Conference on Recent Trends and Directions in Nonparametric Statistics, Part II, 313-328.
- GATHER, U., FRIED, R. (2004): Methods and Algorithms for Robust Filtering, *Proceedings in Computational Statistics*, J. Antoch (Hrsg.), Physica, Heidelberg, 159-170.
- GATHER, U., SCHETTLINGER, K., FRIED, R. (2006): Online Signal Extraction by Robust Linear Regression, *Computational Statistics*, **21**, 35-51.

- IMHOFF, M., KUHL, S. (2006): Alarm Algorithms in Critical Care Monitoring, *Anesthesia and Analgesia*, **102**, 1525-1537.
- JUNG, S.-H. AHN, C. (2000): Estimation of Response Probability in Correlated Binary Data: A New Approach, *Drug Information Journal*, **34**, 599-604.
- KOSKI, E.-M.-J., MÄKIVIRTA A., SUKUVAARA, T., KARI, A. (1990): Frequency and Reliability of Alarms in the Monitoring of Cardiac Postoperative Patients, *International Journal of Clinical Monitoring and Computing*, **7**, 129-133.
- KUHL, S., SIEBIG, S., STÖSSEL, F., IMHOFF, M. (2006): Entwicklung einer Eingabemaske für die Erfassung klinischer Annotationen, SFB 475, Technical Report 15/06.
- LAWLESS, S.-T. (1994): Crying Wolf: False Alarms in a Pediatric Intensive Care Unit, *Critical Care Medicine*, **22**, 981-985.
- LEISENRING, W., PEPE, M.S., LONGTON, G. (1997): Marginal Regression Modeling Framework for Evaluating Medical Diagnostic Tests, *Statistics in Medicine*, **16**, 1263-1281.
- LEE, E., DUBIN, N. (1994): Estimation and Sample Size Considerations for Clustered Binary Responses, *Statistics in Medicine*, **13**, 1241-1252.
- LIANG, K.-Y., ZEGER, S.L. (1986): Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, **73**, 13-22.
- LIU, J.-P., HSUEH, H.-M., HSIEH, E., CHEN, J.J. (2002): Tests for Equivalence or Non-Inferiority for Paired Binary Data, *Statistics in Medicine*, **21**, 231-245.
- LU, Y., JIN, H., GENANT, H.K. (2003): On the Non-Inferiority of a Diagnostic Test Based on Paired Observations, *Statistics in Medicine*, **22**, 3029-3044.
- MÄKIVIRTA, A., KOSKI, E., KARI, A., SUKUVAARA, T. (1991): The Median Filter as a Preprocessor for a Patient Monitor Limit Alarm System in Intensive Care, *Computer Methods and Programs in Biomedicine*, **34**, 139-144.
- MCNEMAR, Q. (1947): Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages, *Psychometrika*, **12**, 153-157.
- NAM, J.-M. (1997): Establishing Equivalence of Two Treatments and Sample Size Requirements in Matched-Pairs Design, *Biometrics*, **53**, 1422-1430.

- NEWCOMBE, R.G. (2001): Simultaneous Comparison of Sensitivity and Specificity of Two Tests in the Paired Design: A Straightforward Graphical Approach, *Statistics in Medicine*, **20**, 907-915.
- OBUCHOWSKI, N.A. (1998): On the Comparison of Correlated Proportions for Clustered Data, *Statistics in Medicine*, **17**, 1495-1507.
- PEPE, M.S. (2003): *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, Oxford.
- RAO, J.N.K., SCOTT, A.J. (1992): A Simple Method for the Analysis of Clustered Binary Data, *Biometrics*, **48**, 577-585.
- SCHETTLINGER, K., FRIED, R., GATHER, U. (2008): Real Time Signal Processing by Adaptive Repeated Median Filters, Preprint.
- SIEGEL, A.F. (1982): Robust Regression Using Repeated Medians, *Biometrika*, **69**, 242-244.
- TANG, M.-L. (2003): Matched-Pair Noninferiority Trials Using Rate Ratio: A Comparison of Current Methods and Sample Size Refinement, *Controlled Clinical Trials*, **24**, 364-377.
- TANGO, T. (1998): Equivalence Test and Confidence Interval for the Difference in Proportions for the Paired-Sample Design, *Statistics in Medicine*, **17**, 891-908.
- TSIEN, C.L., FACKLER, J.C. (1997): Poor Prognosis for Existing Monitors in the Intensive Care Unit, *Critical Care Medicine*, **25**, 614-619.
- WILSON, E.B. (1927): Probable Inference, the Law of Succession, and Statistical Inference, *Journal of the American Statistical Association* **22**, 209-212.
- ZHANG, Y. (2003): *Real-Time Analysis of Physiological Data and Development of Alarm Algorithms for Patient Monitoring in the Intensive Care Unit*, Master Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.
- ZHOU, X.-H., OBUCHOWSKI, N., MCCLISH, D.K. (2002): *Statistical Methods in Diagnostic Medicine*, Wiley Interscience, New York.

Anhang A

Studienprotokoll

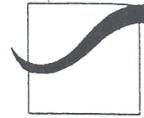


Zentrum für Klinische Studien
Klinikum der Universität Regensburg D-93042 Regensburg

Herrn
Dr. Reng
Klinik und Poliklinik für Innere Medizin I
im Hause

UNIVERSITÄT
REGENSBURG

KLINIKUM



Zentrum für Klinische Studien
Geschäftsstelle der Ethik-Kommission der
Medizinischen Fakultät der Universität

Vorsitzender:
Prof. Dr. med. R. Andreesen
Telefon: +49 - (0)941 - 944-5631
Telefax: +49 - (0)941 - 944-5632
ethikkommission@klinik.uni-regensburg.de

25. Mai. 04

Nr.04/094

Validierung von Alarmalgorithmen in der Intensivmedizin

Der Kommission vorgelegte Unterlagen zur Studie:

Studienprotokoll Version 1.3 vom 05.05.04, Literaturverzeichnis

Sehr geehrter Herr Dr. Reng,

die Ethik-Kommission der Medizinischen Fakultät der Universität Regensburg hat in Ihrer Sitzung am 19.05.2004 über das o.g. Vorhaben auf Grundlage der oben angegebenen Unterlagen beraten. Der Antrag wurde ohne Einschränkungen positiv beschieden.

1. Unabhängig vom Beratungsergebnis verbleibt die ärztliche und juristische Verantwortung beim Projektleiter und seinen Mitarbeitern
2. Die Auflagen der Deklaration von Helsinki des Weltärztebundes in der Fassung von Somerset West / Südafrika 1996 hinsichtlich der ethischen und rechtlichen Aspekte biomedizinischer Forschung am Menschen sind strikt zu beachten.
3. Bitte beachten Sie, dass alle Änderungen oder Erweiterungen in Protokoll sowie der Patienteninformation und -aufklärung schriftlich der Ethik-Kommission mitgeteilt werden müssen. Die Korrekturen sind in der jeweils überarbeiteten Version deutlich kenntlich zu machen.
4. Die Ethik-Kommission erwartet, dass ihr alle schwerwiegenden oder unerwarteten unerwünschten Ereignisse (u.a. alle Todesfälle), die während der Studie auftreten und die Sicherheit der Studienteilnehmer oder die Durchführung der Studie beeinträchtigen können (§40, Abs. 1, Satz 4AMG), unverzüglich schriftlich mitgeteilt werden. Dieses sollte in Verbindung mit einer Stellungnahme des Leiters der Klinischen Prüfung geschehen, ob aus deren Sicht die Nutzen-Risiko-Relation des Vorhabens verändert ist.
5. Der Ethik-Kommission muss der Abbruch oder Abschluss einer Studie mitgeteilt werden. Eine Publikation bzw. der Abschlussbericht der Studie sollte zum gegebenen Zeitpunkt nachgereicht werden. Bei Langzeitstudien bittet die Kommission 12 Monate nach Studienbeginn um Übersendung eines Zwischenberichtes.
6. Handelt es sich bei o.g. Vorhaben um eine Klinische Studie gemäß §§ 19 - 24 MPG, wurde von Seiten der Kommission das Vorliegen aller Voraussetzungen gemäß § 20 Abs. 8 MPG geprüft.

Wir wünschen Ihrem Vorhaben ein gutes Gelingen.

Prof. Dr. R. Andreesen

Hausanschrift:
Zentrum für klinische Studien
Klinikum der Universität Regensburg
Franz-Josef-Strauß-Allee 11
D-93053 Regensburg

Bankverbindung
Sparkasse Regensburg
Konto Nr. 780 010 500
BLZ 750 500 00

ethikkommission@klinik.uni-regensburg.de
<http://www.zks-regensburg.de>

SFB 475 – Studienplan: Validierung von Alarmalgorithmen in der Intensivmedizin

1	Einleitung	2
1.1	Was ist ein Monitor?	2
1.2	Was ist ein Alarm?	2
1.2.1	Ziele von Alarmen	2
1.3	Klinische Situation	3
1.3.1	Häufigkeit von falschen Alarmen	3
1.3.2	Klinische Wahrnehmung falscher Alarme	3
1.3.3	Ursachen von Fehlalarmen	4
1.4	Methodenspektrum	4
1.4.1	Stand der Forschung	4
1.4.2	Eigene Vorarbeiten	6
2	Ansprechpartner	7
3	Bibliographie	7
4	Ziele der Studie	8
5	Dauer der Studie	8
6	Art der Studie	8
7	Studienort	8
8	Patientenkollektiv	8
8.1	Umfang des Patientenkollektivs	8
8.2	Einschlusskriterien	8
8.3	Ausschlusskriterien	9
9	Studiendesign	9
9.1	Datensammlung – Überblick	9
9.2	Patientenidentifikation	9
9.3	Sammlung der Daten vom Patientenmonitor	9
9.4	Sammlung der klinischen Daten durch den Beobachter	10
9.5	Technische Installation	10
9.5.1	Studiengeräte und Zubehör	10
9.6	Dräger-Kontaktperson und Studienmonitor	10
9.7	Unerwartete Nebenwirkungen	11
10	Risikoanalyse	11
11	Patientenaufklärung	11
12	Vertraulichkeit Patienten bezogener Daten	12
13	Unerwünschte Nebenwirkungen, Kodierung und Klassifikation kausaler Abhängigkeiten	12
14	Statistische Methodik und Auswertung	12
14.1	Zielkriterien	12
14.1.1	Hauptzielkriterium	12
14.1.2	Nebenziele	13
14.2	Fragestellung und Hypothesen	13
15	Zusammenfassung	16
16	Literatur	17

1 Einleitung

1.1 Was ist ein Monitor?

Ein Überwachungsgerät, oder auch Monitor, ist ein Werkzeug und keine Therapie. Ein Monitor ist in verschiedenen Aspekten vergleichbar mit einem diagnostischen Gerät. Zahlreiche therapeutische Geräte, z.B. Beatmungsgeräte, besitzen sowohl überwachende als auch therapeutische Komponenten. Eine Anzahl von Attributen wird als charakteristisch für ein Überwachungsgerät angesehen:

- Die Messung von Variablen ist eine intrinsische und essenzielle Eigenschaft.
- Kontinuierliche, automatische Funktion.
- Die Möglichkeit der Warnung durch eine Alarmfunktion.
- Die Möglichkeit, Veränderungen über die Zeit darzustellen.
- Kein direkter therapeutischer Effekt.
- Die Möglichkeit zusätzlicher Datenverarbeitung.

Die Zeitkonstante der Überwachung muss klinisch relevant sein.

1.2 Was ist ein Alarm?

Ein Alarm kann aus klinischer Sicht definiert werden, als eine automatische Warnung resultierend aus der Messung oder sonstigen Erfassung von Variablen (hier: physiologische Variablen) oder sonstigen Deskriptoren eines Zustandes.

In dieser Studie heißen Zustände alarmrelevant, welche eine diagnostische oder therapeutische Entscheidung nach sich ziehen oder welche die Behebung eines technischen Problems, welches die korrekte Datenerfassung dauerhaft beeinträchtigt, erforderlich machen. In der Praxis handelt es sich bei Alarmen meist um Schwellwertalarme, wobei das Über- oder Unterschreiten eines Schwellwertes für eine überwachte Variable zu einer Alarmmeldung in optischer und/oder akustischer Form führt.

1.2.1 Ziele von Alarmen

Aus klinischer Hinsicht sind es verschiedene Ziele, die mit Monitoralarmen erreicht werden sollen. Diese folgen einer bestimmten Hierarchie:

1. Erkennung von lebensbedrohlichen Situationen.
Die Erkennung und Meldung von lebensbedrohlichen Situationen war der Ursprung von Monitoralarmen. Beispiele solcher Situationen umfassen Asystolie, extrem niedrige Blutdrücke oder Hypoxie. In solchen Situationen sind falsch negative Alarme nicht akzeptabel, da sofortige Interventionen notwendig sind, um schweren Schaden oder gar Tod vom Patienten abzuwenden. Dieses Problem ist auch eine Erklärung für die große Häufigkeit falsch positiver Alarme in allen gegenwärtig verfügbaren Monitoren als ein Resultat der notwendigerweise hohen Sensitivität.
2. Erkennung einer drohenden Gefahr.
Die frühzeitige Erkennung einer drohenden Gefahr, z. B. eine fortschreitende Veränderung einer überwachten Variablen, kann Pflegekraft oder Arzt auf eine gefährliche Situation hinweisen, bevor es zur Lebensbedrohung kommt. Eine solche Modalität wurde bisher in keinem kommerziell verfügbaren Monitor implementiert.
3. Diagnostische Alarme.
Ein diagnostischer Alarm könnte auf einen pathophysiologischen Zustand hinweisen, z. B. Hypovolämie, anstatt einfach nur für eine oder mehrere Variablen zu alarmieren. Eine solche Modalität ist heutzutage nicht verfügbar.

Das endgültige Ziel sollte natürlich die frühzeitige Erkennung von Veränderungen sein mit entsprechenden Vorschlägen zu einer adäquaten Therapie. Dies würde eine Online Entscheidungshilfe konstituieren, wie weit über Umfang und Ziel der Patientenüberwachung hinausgeht. Klinische Entscheidungshilfesysteme dieser Art wurden schon im Kontext der klinischen Datenverarbeitung erfolgreich implementiert. Auch wenn gezeigt werden konnte, dass solche Systeme die klinischen Ergebnisse verbessern und Kosten reduzieren können, so ist ihre Entwicklung und Implementierung weiterhin von einem sehr großen Aufwand gekennzeichnet (Übersicht in [48]).

Dies ist nicht Gegenstand der hier beschriebenen Untersuchung und wird auch daher nicht weiter ausgeführt.

1.3 Klinische Situation

1.3.1 Häufigkeit von falschen Alarmen

Es ist eine bekannte Tatsache, dass die große Mehrzahl aller Alarme von Patientenüberwachungsgeräten, unabhängig von Monitorart und Hersteller, falsch positive Alarme sind.

In der medizinischen Literatur gibt es zahlreiche Studien über die Häufigkeit von falschen Alarmen in der Intensivüberwachung. Diese Untersuchungen zeigen, dass bis zu 90% aller Alarme in der intensivmedizinischen Überwachung falsch positiv sind [50]. In vielen Fällen resultieren sie aus Mess- oder Bewegungsartefakten. Die überwiegende Mehrheit aller Schwellwertalarme auf der Intensivstation ist ohne klinische Relevanz für die Behandlung des kritisch Kranken [40, 44]. In einer Studie resultierten 75 Prozent aller Alarme in keiner medizinischen oder pflegerischen Handlung [11]. Bei dieser Untersuchung lagen der positive prädiktive Wert der Alarme nur bei 27%, die Spezifität nur bei 58%. Der negative prädiktive Wert und die Sensitivität lagen bei 99% bzw. 97%.

Diese Ergebnisse werden unterstützt von früheren Studien, die den Anteil klinisch relevanter Alarme an allen Alarmen mit 10% [37] oder 5% [40] beziffern. Während 27% induzierte Alarme waren, d. h. verursacht durch pflegerische Maßnahmen, waren 68% tatsächlich falsche Alarme. Der positive prädiktive Wert lag zwischen 5% und 16% in Abhängigkeit von der überwachten Variablen.

Bisher konnten nur punktuelle Verbesserungen erzielt werden, z.B. durch verbesserte digitale Arrhythmieerkennung oder neue Hardware und Software zur Pulsoxymetrie.

1.3.2 Klinische Wahrnehmung falscher Alarme

Die Häufigkeit falscher Alarme führt zu einer gefährlichen Desensibilisierung der Mitarbeiter auf Intensivstationen gegenüber wahren Alarmen [47]. Es gibt nicht nur zu viele Alarme, von denen viele falsch sind, sondern es gibt auch viele verschiedene Alarme oder unterschiedliche akustische oder optische Meldungen für die gleiche klinische Situation. In einer Studie konnten Intensivpflegekräfte nur 38% aller akustischen Alarme korrekt identifizieren [13]. Eines der klinischen Resultate dieser schlechten Alarmqualität wurde in einer anderen aktuellen klinischen Studie gezeigt [38]: 23 Kardioanästhesisten und Kardiochirurgen wurden nach ihrer Meinung gefragt, wie die Monitoralarme klinisch adäquat einzustellen seien. Es wurden nur Herzfrequenz- und Blutdruckalarme genannt, während alle anderen Alarme deaktiviert werden sollten. Darüber hinaus wurden die Alarmgrenzen soweit gesetzt, dass sogar lebensbedrohliche Situationen nicht in einem Alarm resultieren würden. Selbst unter diesen Einstellungen würden die Kliniker einen Alarm bis zu 10 Minuten lang tolerieren, bevor sie einschreiten würden.

Aus unseren klinischen Erfahrungen und von den zitierten Studien ist es offensichtlich, dass die gegenwärtige Qualität von Alarmen in der Intensivmedizin nicht akzeptabel und sogar

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

gefährlich für den Patienten ist. Da falsche Alarmer ein allgegenwärtiges Problem mit jeder Art des physiologischen Monitorings ist, ist es sehr erstrebenswert, Methoden zu entwickeln, die automatisch klinisch relevante Alarmsituationen erkennen und melden können.

1.3.3 Ursachen von Fehlalarmen

Falsche Alarmer können verursacht werden durch technische Probleme, wie z. B. die falsche Wahl der EKG Ableitung, eine zu niedrige EKG Spannung, Dämpfung von Druckleitungen, usw., oder durch Artefakte, wie Patientenbewegungen, das Spülen von Druckleitungen, Blutabnahmen, usw. Ein dritter Grund für falsche Alarmer ist die klinische Unangemessenheit des Alarms selbst, der Alarmgrenzen oder der Algorithmen.

In einer Studie stammte ungefähr ein Drittel aller Alarmer vom Beatmungsgerät, ein weiteres Drittel von der Kreislaufüberwachung und ca. 15% von der Pulsoximetrie allein [11]. Eine andere Studie fand, dass die Pulsoximetrie für über 40% aller falschen Alarmer verantwortlich war [40]. Auch wenn die Pulsoximetrie durch neue Algorithmen und neue Sensoren deutlich verbessert werden konnte, ist sie weiterhin neben dem EKG Monitoring die führende Quelle falscher Alarmer.

Somit ist offensichtlich, dass die gegenwärtige Alarmgebung in der Akut- und Intensivmedizin nicht zufrieden stellend ist. Es stellt sich daher die Frage, inwieweit statistische Methoden geeignet sind, diese Situation zu verbessern.

1.4 Methodenspektrum

1.4.1 Stand der Forschung

Die Analyse von Online-Monitoring-Daten ist seit SHEWHART (1931) fest verankert in der statistischen Prozesskontrolle mit dem Ziel, Industrieprozesse in Echt-Zeit zu überwachen [58]. Der erste Ansatz zur Erkennung von Change-Points bei medizinischen Daten geht auf TRIGG (1964) zurück [63]. Ein Überblick über weitere frühere Verfahrensweisen findet sich bei HILL und ENDRESEN (1978) [30].

Häufig auftretende Charakteristika intensivmedizinischer Zeitreihen sind große und veränderliche Prozessvariabilität, Instationarität, Messfehler, das Auftreten von Ausreißern sowie Niveau- und Trendänderungen. Zeitweise ist die Datenqualität eingeschränkt, bedingt durch labilen Patientenzustand sowie pflegerische und medizinische Maßnahmen; ebenso sind fehlende Messwerte zu erwarten.

Eine statistische Methodik zur Kontrolle intensivmedizinischer Variablen muss in der Lage sein, Ausreißer, Niveauänderungen und Trends in den Zeitreihen zu erkennen und dies, möglicherweise unter Ausnutzung der Abhängigkeitsstruktur in den multivariaten Zeitreihen, bevor der Patientenzustand kritisch wird. Eine weitere grundlegende Anforderung ist die Notwendigkeit eines individuellen Patientenmonitorings.

1.4.1.1 Zeitreihenanalysen in der Intensivmedizin

Zeitreihenmethoden im allgemeinen und ARIMA-Modelle im besonderen [6] fanden in großem Umfang in psychologischen, psychometrischen, soziologischen, epidemiologischen und weiteren Untersuchungen Verwendung [4, 7, 9, 28, 51, 52, 56, 57, 62, 65, 66, 67]. Im Gegensatz hierzu wurden bislang nur wenige Untersuchungen in der Intensivmedizin durchgeführt [36]. Dies ist um so überraschender, als in der Intensivmedizin große Datenmengen in unterschiedlichen Zeitintervallen von einzelnen oder mehreren Patienten erhoben, registriert und gespeichert werden. Gerade Veränderungen von Messwerten im Zeitverlauf als Folge von pathophysiologischen Prozessen oder therapeutischen Interventionen bieten ein weites Anwendungsfeld für Zeitreihenmethoden. Im Prinzip können

Version: 1.3	Letzte Änderung: 5.05.04	Seite 4 von 19
--------------	--------------------------	----------------

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

nur mit derartigen statistischen Werkzeugen sogenannte Einzelfallstudien sinnvoll durchgeführt werden [12, 21, 31, 59].

So konnte gezeigt werden, dass Zeitreihenanalysemethoden für die retrospektive Untersuchung von physiologischen Variablen geeignet sind [32, 33, 39]. In der Literatur finden sich autoregressive Modelle, die in longitudinalen physiologischen Experimenten [39] sowie bei der Analyse von Labordaten von chronisch Kranken [41]. Analysen von Therapieeffekten mit autoregressiven Modellen wurden aus der Kardiologie [39], der Neurologie [20], der Psychiatrie [45], sowie der Pflegeforschung [29] berichtet.

Eine statistisch vergleichbare, aber methodisch deutlich anspruchsvollere Aufgabe ist die Analyse von intensivmedizinischen Monitoringdaten in Echtzeit. Von den wenigen zur Zeit bekannten statistischen Lösungsansätzen ist keiner direkt in die klinische Anwendung übertragbar [21, 60, 61].

Zur Überwachung des Serum-Kreatininwertes bei Patienten, die einer Nierentransplantation unterzogen wurden, setzen SMITH und Mitarbeiter erstmals einen Multi-Prozess-Kalman-Filter ein [60, 61]. Diese Methodik nutzt das lineare dynamische Multi-Prozess-Modell in Verbindung mit einem Bayes-Ansatz. Andere Arbeiten zur Modellierung biomedizinischer Zeitreihen mit Multi-Prozess-Modellen finden sich bei GORDON sowie GORDON und SMITH [21, 22].

In den letzten Jahren wurden verstärkt ARIMA-Modelle für die Analyse intensivmedizinischer Daten herangezogen [4, 32]. Diese Analysen wurden retrospektiv durchgeführt, um zunächst die Eignung der Modelle für die Beschreibung intensivmedizinischer Daten zu prüfen. Dabei wurde die Verwendbarkeit insbesondere von AR-Modellen niedriger Ordnung bei hämodynamischen Variablen gezeigt. Darüberhinaus wurden AR-Modelle mit additiven Interventionsvariablen zur Beurteilung des Effektes von Beatmungstherapien wie IRV und PEEP bei Patienten mit akutem Lungenversagen herangezogen [32].

Diese Arbeiten zur Modellierung des Steady State mit ARIMA-Modellen zeigen, dass hiermit eine verbesserte und objektivere Verlaufsbeurteilung von physiologischen Variablen und Beurteilung von Interventionseffekten gegenüber der rein klinischen Bewertung möglich ist [32, 33, 36]. Auch die Anwendbarkeit von AR-Modellen zur Erkennung und Unterscheidung von Mustern wie Ausreißern, Leveländerungen und Trends in physiologischen Daten wurde von IMHOFF, BAUER, GATHER et al. untersucht [33, 34].

1.4.1.2 Zeitreihenanalysen im Vergleich zu anderen Analyseverfahren

Für die Alarmverarbeitung wurden von MÁKIVIRTA et al. [44] Medianfilter eingesetzt. Diese eignen sich hervorragend, um Rauschen und vor allem punktuelle Artefakte vor der eigentlichen Analyse zu entfernen. Allerdings werden Autokorrelationen nicht berücksichtigt und klinisch relevante Muster, wie Niveauänderungen oder Trends, können allein hiermit nicht unterschieden werden.

Methoden der Kurvenanpassung wurden ebenfalls eingesetzt, um Zeitreihendaten zu analysieren. HAIMOWITZ et al. [24, 25] entwickelten das Program TrenDx um klinisch relevante Trends im kindlichen Wachstumsverhalten zu entdecken. Hier wurden die beobachteten Zeitreihen mit vorher definierten Basisfunktionen verglichen. Für eine präzise Analyse müssen dementsprechend zahlreiche Basisfunktionen definiert werden. Zudem bleiben auch bei diesem Ansatz Autokorrelationen unberücksichtigt.

Bei industriellen Kontrollkarten, wie sie z.B. durch SHEWHART [58] eingeführt wurden, wird von einem festen Basiswert ausgegangen. Dies ist bei physiologischen Daten nicht unproblematisch. Zudem wird im klassischen Ansatz nicht zwischen verschiedenen Mustern unterschieden, was aber gerade für die klinisch wichtige Unterscheidung zwischen Artefakten und medizinisch relevanten Mustern von hohem Interesse ist.

Version: 1.3	Letzte Änderung: 5.05.04	Seite 5 von 19
--------------	--------------------------	----------------

1.4.2 Eigene Vorarbeiten

1.4.2.1 Methodik und Vorgehensweise

Einen Schwerpunkt der bisherigen Arbeiten stellen Methoden zur Online-Mustererkennung in univariaten Zeitreihen dar. Zur Erkennung und Unterscheidung von Ausreißern und Niveauänderungen in univariaten Zeitreihen schlägt Bauer [2] eine Phasenraumrekonstruktion für lineare stationäre Zeitreihen vor. Auf dieser Darstellung basierend wurde eine automatische Prozedur entwickelt, die als eine Erweiterung der Shewhart-Karte auf abhängige Daten angesehen werden kann.

Diese Phasenraumkarte wird durch eine Online-Methode zur Trenderkennung ergänzt. Hierzu wurden zum einen Ansätze basierend auf exponentiellem Glätten untersucht. Zum anderen wurde der Mini-Max-Ansatz von Abelson und Tukey [1] und Brillinger [8] basierend auf der Technik gleitender Zeitfenster für die Online-Anwendung weiterentwickelt. Für die notwendige Autokovarianzschätzung wurden stückweise lineare Mittelwertanpassungen verschiedener Ordnung zur lokalen Trendbereinigung implementiert [17].

Viele Zeitreihenanalyseverfahren wie etwa ARMA-Modelle, aber auch die Phasenraumkarte oder Kontrollkarten für dynamische Daten erfordern eine Schätzung der Autokovarianzfunktion. Da klassische Schätzer stark von Ausreißern beeinflusst werden können, wurden mehrere robuste Alternativen anhand von Simulationen in verschiedenen Datensituationen verglichen [10].

Die robuste Extraktion eines zeitveränderlichen Mittelwertes aus physiologischen Zeitreihen ist von elementarer Bedeutung bei der Beurteilung des Patientenzustandes und kann als Vorstufe zur Modellierung instationärer Zeitreihen dienen. Dazu wurden robuste Regressionsverfahren wie die L1-Regression, der Least Median of Squares [26, 53] und der Repeated Median [59] zur Approximation eines lokal-linearen Trends mittels eines gleitenden Zeitfensters in grundlegenden Datensituationen verglichen. Um die Robustheit dieser Methoden weiter zu verbessern wurden automatische Ausreißerersatzregeln wie Trimming und Winsorisieren basierend auf robusten Skalenschätzern wie dem klassischen MAD, der Length of the Shortest Half [23, 55] und der Q_α -Methode [54] eingebaut.

1.4.2.2 Ergebnisse und Einordnung

Das vordringliche Problem der Online-Mustererkennung in univariaten Zeitreihen kann als weitgehend gelöst gelten. Eine umfangreiche Studie zeigt, dass AR(2)-Modelle eine deutliche Verringerung der Fehlalarmrate gegenüber AR(0)- und AR(1)-Modellen bei gleicher Zuverlässigkeit (Sensitivität) ermöglichen, während die Unterschiede zu AR-Modellen höherer Ordnung klinisch nicht relevant sind [35]. Insgesamt liefern AR-Modelle zweiter Ordnung für Zeitreihen von Herzkreislaufvariablen mit einem Erhebungsintervall von einer Minute eine zuverlässige Mustererkennung bei kurzer Rechenzeit. Eine Untersuchung des Einflusses der Länge des Schätzintervalls auf die Mustererkennung zeigt die negativen Auswirkungen von Strukturbrüchen im Schätzintervall und damit die Notwendigkeit robuster Methoden auf.

Eine Vergleichsstudie belegt die Überlegenheit des Phasenraumansatzes bei der Erkennung von Ausreißermustern gegenüber Vorhersage-basierten Ansätzen [18]. Weitere Verbesserungen konnten durch eine robuste Autokovarianz-Schätzung mittels MCD-Schätzern erzielt werden.

Die Online-Erkennung langsamer Trends basierend auf exponentieller Glättung hängt entscheidend vom Glättungsparameter ab, der in einer Größenordnung von 0.05 oder kleiner gewählt werden sollte [43]. Für die entwickelte Online-Variante des Mini-Max-Ansatzes von Abelson und Tukey [1] ist stattdessen eine geeignete Fensterlänge zu wählen, was intensivmedizinischen Anforderungen wie etwa zuverlässiger Erkennung von über 30 bis 60

Minuten anhaltenden monotonen Trends jeglicher Form entgegenkommt. Zur Schätzung der Autokorrelationsfunktion erweist sich die Anpassung einer Mittelwertfunktion relativ hoher Ordnung zur Trendbereinigung unter Einsatz der Shrinkage-Methode zur Vermeidung einer Überspezifikation als empfehlenswert. Die Anwendung auf simulierte und echte Zeitreihen zeigt, dass lineare und nichtlineare Trends im Falle realistischer moderat positiver Autokorrelationen zuverlässig und ohne viele Fehlalarme erkannt werden können [17]. Bei sehr hohen positiven Autokorrelationen ist wie beim exponentiellen Glätten eine große Beobachtungsanzahl und damit ein starker Zeitverzug zur Unterscheidung systematischer Trends von positiven Autokorrelationen erforderlich.

Auch die robuste Echt-Zeit Signalextraktion aus univariaten Zeitreihen kann als gelöst angesehen werden. Als Mittel der Wahl ist hier der Repeated Median hervorzuheben, der in Trendperioden eine höhere Effizienz und Robustheit als der klassische Median aufweist [16]. Zwar zeigt der Least Median of Squares Schätzer (LMS) eine größere Robustheit bei einem hohen Prozentsatz von Ausreißern, doch neigt er zu Instabilitäten bei kleinen Änderungen in den Daten und benötigt deutlich höhere Rechenzeiten [14]. Der Kontaminationsbias des Repeated Median konnte durch Verwendung automatischer Ausreißerersetzungsregeln, insbesondere Trimming, erheblich verbessert werden. Hierbei erwiesen sich zur Skalenschätzung hauptsächlich die Q_α -Methode und die Length of the Shortest Half (LSH) als gut geeignet [19]. Während die Q_α -Methode zu sehr guten Ergebnissen bei so genannten Inliers und bei Ausreißern vergleichbarer Größenordnung, etwa in Folge einer Leveländerung, führt, besitzt LSH ein besseres Worst-Case Verhalten. Die automatische Ersetzung von Ausreißern bewirkt eine zum LMS vergleichbar hohe Robustheit des Repeated Median [15]. Rechenzeiten bedeuten für den Repeated Median auch im Online-Einsatz kein Hindernis, zumal ein Algorithmus für den Update der Schätzung beim Bewegen des Zeitfensters entwickelt werden konnte, der nur $O(n)$ statt $O(n^2)$ Zeit im Vergleich zu einer Neuberechnung erfordert [5].

Insgesamt lässt sich festhalten, dass Methoden der statistischen Zeitreihenanalyse sehr gut geeignet sind, klinisch relevante Muster in univariaten Zeitreihen aus dem Intensivmonitoring in Echtzeit zu erkennen und von nicht klinisch relevanten, wie z.B. Artefakten, zu unterscheiden. Dieses konnte in Studien mit umfangreichen klinischen Daten nachgewiesen werden [17, 18, 19, 34, 35]. In Verbindung mit der sehr effizienten algorithmischen Umsetzung dieser Verfahren erscheinen sie besonders geeignet zum Einsatz in der Alarmgebung im intensivmedizinischen Online Monitoring.

2 Ansprechpartner

Die Verantwortlichkeiten für diese klinische Studie sind wie folgt verteilt:

- Klinischer Projektleiter: Dr. Michael Reng.
- Stellvertretende klinischer Projektleiter: Dr. Christian Wrede, Dr. Felix Rockmann
- Leiter der klinischen Prüfung: Dr. Michael Reng.
- Statistische Projektleitung: Prof. Dr. U. Gather, Dipl.-Stat. Vivian Lanus, PD Dr. Michael Imhoff.
- Studienmonitor: Christina De Mur, Draeger Medical Systems, Inc. (Danvers, Massachusetts).
- Studiensponsor: Draeger Medical Systems, Inc. (Danvers, Massachusetts), Deutsche Forschungsgemeinschaft (SFB 475).

3 Bibliographie

1. Food and Drug Administration Regulation; 21 CFR Parts 50 and 812.25
2. ISO Clinical Trial Document 14155: 1996

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

3. EN 540,1993
4. Declaration of Helsinki: Recommendation for Conduct of Clinical Research
5. Bedienungsanleitungen der verwendeten medizintechnischen Geräte.

4 Ziele der Studie

Hauptziel der Studie ist die klinische Validierung neuer Algorithmen zur Alarmgebung im Intensivmonitoring. Überprüft werden soll, ob mit neu entwickelten statistischen Online-Verfahren zur Alarmgebung die Fehlalarmrate bei unvermindert hoher *Sensitivität* gegenüber den in der Praxis eingesetzten Schwellwertalarmen reduziert werden kann.

Zusätzlich werden folgende Nebenziele verfolgt:

1. Die Erstellung eines Referenzdatensatzes durch Sammlung sämtlicher relevanten Patientendaten inklusive einer möglichst zeitnahen klinischen Bewertung von Alarmsituationen. Dieser Datensatz soll bei der Evaluation und Entwicklung der hier untersuchten und zukünftiger Alarmalgorithmen eingesetzt werden können.
2. Die Untersuchung der Häufigkeit von Alarmen und der klinischen Relevanz/Validität der existierenden Alarmalgorithmen.
3. Es soll überprüft werden, ob mit neu entwickelten statistischen Verfahren zur Online-Mustererkennung bei geringer Fehlalarmrate und hoher Sensitivität Frühwarn-Alarme möglich sind.
4. Die Anwendung und Untersuchung von Methoden der Dimensionsreduktion.

5 Dauer der Studie

Die Studie ist auf einen Zeitraum von zunächst 18 Monaten ausgelegt. Die Studienzeit kann verlängert werden, um die notwendige Zahl von Patienten einzuschließen.

6 Art der Studie

Offene Beobachtungsstudie.

7 Studienort

Die Studie wird durchgeführt auf der Intensivstation der Klinik und Poliklinik für Innere Medizin I des Klinikums der Universität Regensburg.

8 Patientenkollektiv

8.1 Umfang des Patientenkollektivs

In die Studie sollen mindestens 100 Patienten eingeschlossen werden.

Die Häufigkeit von Alarmen und deren zugrunde liegende Zustände können nicht verlässlich vorhergesagt werden. Daher kann auch die Zahl der Patienten, die mindestens eingeschlossen werden sollen, nicht vorab festgelegt werden. Das Ziel der Studie ist es, so viele Daten zu sammeln, wie in der zur Verfügung stehenden Zeit möglich ist. Es bestehen keine Präferenzen bezüglich Alter, Geschlecht oder Erkrankungsschwere.

8.2 Einschlusskriterien

Patienten, die folgende Kriterien erfüllen, sollen in die Studie aufgenommen werden:

- Patienten älter als 18 Jahre, die mit einem Dräger Infinity Monitor überwacht werden, und bei denen mindestens kontinuierlich Herzfrequenz, kontinuierlich invasiv oder

Version: 1.3	Letzte Änderung: 5.05.04	Seite 8 von 19
--------------	--------------------------	----------------

SFB 475

Studienplan: Alarmstudie

kontinuierlich/quasi kontinuierlich nicht-invasiv arterieller Blutdruck und kontinuierlich Sauerstoffsättigung gemessen werden.

- Bevorzugt werden Patienten, bei denen möglichst viele Herzkreislaufvariablen überwacht werden, sowie Patienten, bei denen ein häufiges Vorkommen von Alarmen zu erwarten ist.

8.3 Ausschlusskriterien

Folgende Patienten werden von der Studie ausgeschlossen:

- Patienten, die das 18. Lebensjahr nicht vollendet haben.
- Patienten, bei denen nur eine Variable, z.B. Herzfrequenz gemessen wird.
- Patienten, bei denen der Blutdruck nicht invasiv gemessen wird.
- Patienten, bei denen alle Alarmer über einen längeren Zeitraum inaktiviert sind.

9 Studiendesign

9.1 Datensammlung – Überblick

Das Ziel der elektronischen Datensammlung ist es folgende Daten zu erfassen:

1. Alle numerischen Messwerte des jeweiligen Patientenmonitors.
2. Alle Messwert-Echtzeitkurven (in Abhängigkeit von der technischen Realisierbarkeit).
3. Alle Alarmmeldungen.
4. Alle Alarmerstellungen.
5. Klinische Annotationen zur Bewertung der Alarmsituationen.
6. Andere relevante Patientendaten.
7. Regelmäßige Bewertung des Patientenzustands, z.B. in den Kategorien kritisch, abwärts, schwankend, aufwärts, stabil, in klinisch hinreichend engen Intervallen.

Alle automatisch zu erfassenden Daten, also alle Daten vom Patientenmonitor, werden kontinuierlich in einer speziellen Datenbank erfasst (Pos. 1.-4.). Diese Datensammlung erfolgt während der Beobachtungszeit durch den klinischen Beobachter. Zusätzlich können alle Monitordaten auch kontinuierlich über einen längeren Zeitraum ohne klinische Beobachtungen erfasst werden.

Zusätzlich werden von einem geschulten Beobachter die auftretenden Alarmsituationen, d.h. Situationen, in denen der Monitor einen Alarm aussendet, klinisch bewertet (Pos 5.). Daneben werden auch Umstände erfasst, die das Auftreten eines Alarms wahrscheinlich machen, z.B. Bewegungen des Patienten (Pos. 6.). Der Beobachter greift nicht in den Pflege- und Therapieprozess ein. Der Beobachter wird jeweils 8 Stunden kontinuierlich einen oder zwei Patienten (in einem Zwei-Bett-Zimmer) beobachten.

9.2 Patientenidentifikation

Allen Patienten wird im Monitorsystem eine fortlaufende, eindeutige Identifikationsnummer automatisch zugewiesen. Diese Identifikationsnummer wird auch zur Identifikation der Patienten verwendet, bei denen die klinischen Daten manuell erfasst werden. Anhand dieser Identifikationsnummer werden zur Analyse die entsprechenden Datensätze zusammengeführt. Den Patienten mit klinischer Datenerfassung wird zusätzlich eine laufende Studiennummer zugeordnet.

9.3 Sammlung der Daten vom Patientenmonitor

Alle Monitordaten, wie in Kapitel 9.1 (Punkte 1-4) aufgelistet, werden kontinuierlich von allen Monitoren, die auf der Intensivstation der Inneren Abteilung in Dräger INFINITY®-Netzwerk angeschlossen sind, von einem spezielle Netzwerkserver (Dräger GATEWAY®

Version: 1.3

Letzte Änderung: 5.05.04

Seite 9 von 19

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

Server) auf einen zweiten dedizierten Studienrechner weitergeleitet. Hier werden die Daten entsprechend gespeichert.

Eine genaue Beschreibung des Datenexportprotokolls, der Netzwerkkonfiguration sowie der Datenspeicherung wird im Anhang gegeben.

Eine Beeinflussung des Monitornetzwerkes oder einzelner Monitore ist durch dieses Vorgehen vollkommen ausgeschlossen. Bei dem GATEWAY® Server handelt es sich um ein zertifiziertes medizintechnisches Produkt, dessen primärer Verwendungszweck der Export von Monitordaten aus dem INFINITY®-Netzwerk in ein anderes, offenes Netzwerk ist.

9.4 Sammlung der klinischen Daten durch den Beobachter

Der klinische Beobachter erfasst mit Hilfe des Studiennotebooks und einer entsprechenden Eingabemaske folgende Beobachtungen:

- Bei Auftreten einer Alarmsituation
 - Validität des Alarms
 - Löschen des Alarms
 - Reaktion des Pflegepersonals
 - Vermutliche klinische Ursache des Alarms
- Außerhalb einer Alarmsituation
 - Klinische Umstände, die zu einem Alarm führen können.
 - Sonstige klinische Beobachtungen.

Vor Beginn jeder Datensammlung wird die Systemuhr des Notebooks mit der Systemuhr des Monitorsystems synchronisiert auf +/- 30sec. Dabei ist die Monitorsystemuhr der Zeitmaster. Jedem neuen Patienten wird die Identifikationsnummer zugewiesen, die er vom Monitorsystem erhalten hat. Zusätzlich wird eine fortlaufende Studiennummer zugewiesen. Das Vorgehen ist weiter unten detailliert beschrieben.

Die genaue technische und softwaregemäße Umsetzung ist im Anhang zu einem späteren Zeitpunkt beschrieben.

9.5 Technische Installation

Alle mit den Patienten verbundenen Geräte sind für die jeweilige Anwendung CE- und FDA-zertifiziert.

9.5.1 Studiengeräte und Zubehör

Folgende Geräte werden im Rahmen der Studie eingesetzt:

Beschreibung	Anzahl	Kommentar
Dräger Infinity Monitoringsystem	1	Auf den Intensivstationen vollständig vorhanden mit einem Monitor je Bettplatz.
Dräger GATEWAY®-Server	1	Im Klinikum vorhanden
PC zur Datensammlung vom GATEWAY®-Server	1	Genaue Ausstattung TBD. Netzwerkanbindung TBD.
Notebook mit Software zur Datenakquisition	1	Endgültige SW-Ausstattung TBD.

9.6 Dräger-Kontaktperson und Studienmonitor

Die zuständige Kontaktperson bei Dräger Medical Systems und zugleich Studienmonitor ist: Christina De Mur

Sr. Clinical Product Manager

Version: 1.3	Letzte Änderung: 5.05.04	Seite 10 von 19
--------------	--------------------------	-----------------

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

Draeger Medical Solutions, Inc.
16 Electronics Ave
Danvers, MA 01923
+1-978-907-7691
Email: christina.demur@draeger.com

9.7 Unerwartete Nebenwirkungen

Jede(s) unerwartete/unerwünschte Ereignis/Nebenwirkung in Verbindung mit der Benutzung der Geräte für diese Studie muss innerhalb von 24 Stunden nach Auftreten des Ereignisses telefonisch und innerhalb von 48 Stunden schriftlich mitgeteilt werden an:

Connie Hertel
Director Quality Assurance and Management
Draeger Medical Solutions
16 Electronics Ave.
Danvers, MA 01923
+1-978-907-7500
Email: connie.hertel@draeger.com

Zudem werden die zuständige Ethikkommission, welche die Zustimmung zu dieser Studie erteilt, und die Food and Drug Administration informiert im Falle eines unerwarteten/unerwünschten Ereignis/Nebenwirkung.

10 Risikoanalyse

Die verwendeten Geräte bieten keinerlei elektrische Risiken für die Patienten. Sie sind alle konform mit den elektrischen Sicherheitsstandards UL 544 und IEC 601-1.

Die Antragsteller und Dräger Medical Systems, Inc., vertreten die Auffassung, dass diese klinische Studie in die Kategorie „non-significant risk“ fällt entsprechend der Definition in 21 CFR 812.2 part b. In diesem Sinne wird angenommen, dass diese Untersuchung in keinerlei zusätzlichem Risiko für die Patienten resultiert und daher in Übereinstimmung mit 21 CFR 56.109 als „minimal risk investigation“ eingestuft werden soll.

In der vorliegenden Studie werden Daten nur passiv vom Monitor gesammelt zur weitergehenden Analyse außerhalb des Behandlungsbereiches. Es besteht keinerlei Notwendigkeit zusätzlicher invasiver oder nicht-invasiver Prozeduren beim Patienten. Modifikationen an Software oder Hardware der verwendeten Patientenmonitore werden nicht vorgenommen, so dass keinerlei Beeinflussung von Diagnostik, Überwachung oder Therapie der Patienten oder eine Interferenz mit der ärztlichen oder pflegerischen Tätigkeit vorstellbar sind.

Die während der Studie erhobenen Daten werden zu keinem Zeitpunkt zur Beeinflussung medizinischer (diagnostischer, therapeutischer) Entscheidungen für die Studienpatienten herangezogen.

Es besteht somit keinerlei zusätzliches Risiko für den Patienten durch die vorliegende Studie. Ein Vorteil für den Studienpatienten besteht ebenfalls nicht.

11 Patientenaufklärung

Eine Patientenaufklärung wird für nicht notwendig erachtet, da keinerlei neue Verfahren oder Techniken am Patienten getestet werden, eine Erweiterung diagnostischer oder therapeutischer Interventionen durch die Studie nicht vorgenommen wird und die Erfassung der Patientenidentität oder ein Follow-up für die Studie nicht erforderlich sind.

Version: 1.3	Letzte Änderung: 5.05.04	Seite 11 von 19
--------------	--------------------------	-----------------

SFB 475

Studienplan: Alarmstudie

12 Vertraulichkeit Patienten bezogener Daten

Alle Daten werden nur unter einer fortlaufenden Patientennummer gespeichert. Daten, die der Identifikation des Patienten dienen könnten, werden nicht gespeichert. Eine Zuordnung von Patientennummer und tatsächlicher Patientenidentität ist nur dem Prüfarzt möglich.

13 Unerwünschte Nebenwirkungen, Kodierung und Klassifikation kausaler Abhängigkeiten

Da keine Interventionen am Patienten erfolgen, sind unerwünschte, der Studie unmittelbar oder mittelbar zuzuordnende Nebenwirkungen nicht zu erwarten, und damit erscheint eine Kodierung/Klassifikation nicht sinnvoll. Sollten dennoch während der Studie nicht vorhersehbare Ereignisse auftreten, die eine mittelbare oder unmittelbare Gefährdung des Patienten *möglich* erscheinen lassen, werden diese Ereignisse/Gegebenheiten sowohl der Studienzentrale, dem Studienmonitor und den leitenden Studienärzten als auch der Ethikkommission zeitnah mitgeteilt.

14 Statistische Methodik und Auswertung

14.1 Zielkriterien

14.1.1 Hauptzielkriterium

Das **Hauptzielkriterium** ist die klinische Validität von Alarmalgorithmen im Intensivmonitoring.

Überprüft werden soll, ob mit neu entwickelten statistischen Online-Verfahren zur Alarmgebung die Fehlalarmrate bei unvermindert hoher *Sensitivität* gegenüber den in der Praxis eingesetzten Schwellwertalarmen reduziert werden kann.

Dabei bezieht Sensitivität sich auf das *Erkennen alarmrelevanter Zustände*. Ein Zustand heißt *alarmrelevant*, wenn er eine diagnostische oder therapeutische Entscheidung nach sich zieht oder wenn er die Behebung eines technischen Problems, welches die korrekte Datenerfassung dauerhaft beeinträchtigt, erforderlich macht. Diese Definition beinhaltet keine Wertung in Bezug auf den klinischen Zustand des Patienten oder den Schweregrad dieses Zustandes sondern orientiert sich nur am Handlungsspektrum. In der Literatur zu anderen Studien zu Alarmsystemen in der Intensivmedizin werden alarmrelevante Zustände ähnlich definiert (siehe CHAMBRIN et al. (1999), KOSKI et al. (1990), LAWLESS (1994) und TSIEN (2000) [11, 37, 40, 64]).

Zur *Erkennung* der alarmrelevanten Zustände sollen robuste Online-Signalextraktionsverfahren verbunden mit Regeln zur Ausreißererkennung und Schwellwertüberschreitung zum Einsatz kommen.

Alarmrelevante Zustände können mit Hilfe von Annotationen, die während der Datenerhebung von einem speziell ausgebildeten Beobachter gemacht werden, identifiziert werden. Jeder Alarm, der von einem auf Intensivstationen eingesetzten Patientenüberwachungssystem ausgelöst wird, wird von dem Beobachter

- als *richtig* bewertet, wenn ein alarmrelevanter Zustand gemäß obiger Definition vorliegt;
- als *herbeigeführt* eingestuft, wenn er durch Intervention des Personals ausgelöst wurde und der Zustand nicht alarmrelevant ist;
- als *falsch* klassifiziert, wenn der Alarm nicht herbeigeführt wurde und kein alarmrelevanter Zustand vorliegt.

Zudem wird in den Annotationen festgehalten, wann ein wie oben definierter alarmrelevanter Zustand vorlag, ohne dass vorher ein Alarm ausgelöst wurde.

Version: 1.3

Letzte Änderung: 5.05.04

Seite 12 von 19

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

Mit Hilfe solcher Annotationen kann nicht nur das zurzeit auf Intensivstationen eingesetzte Schwellwertalarmsystem, sondern jedes auf neuen Algorithmen beruhende System beurteilt werden.

14.1.2 Nebenziele

Die Studie hat folgende **Nebenziele**:

1. Die Erstellung eines Referenzdatensatzes durch Sammlung sämtlicher relevanten Patientendaten inklusive einer möglichst zeitnahen klinischen Bewertung von Alarmsituationen. Dieser Datensatz soll bei der Evaluation und Entwicklung der hier untersuchten und zukünftiger Algorithmen eingesetzt werden können.
2. Die Untersuchung der Häufigkeit von Alarmen und der klinischen Relevanz/Validität der existierenden Algorithmen.
3. Es soll überprüft werden, ob mit den neu entwickelten statistischen Verfahren zur Online-Mustererkennung bei geringer Fehlalarmrate und hoher Sensitivität die Auslösung von Frühwarn-Alarmen möglich ist. Die neuen Algorithmen zur Mustererkennung umfassen Prozeduren zur Erkennung von Ausreißern und Niveauänderungen, zur Erkennung von Trends und Regeln, die auf robuster Online-Signalextraktion basieren. Außerdem wird eine Kombination dieser Verfahren betrachtet.
4. Die Anwendung und Untersuchung von Methoden der Dimensionsreduktion.

14.2 Fragestellung und Hypothesen

Die durch das Hauptzielkriterium formulierte Aufgabe besteht in dem Vergleich zweier Alarmsysteme zur kontinuierlichen Überwachung der Stabilität des Zustands von Patienten auf der Intensivstation. Dieses Problem kann auf den Vergleich der Güte zweier Diagnoseverfahren bei vorhandener Referenz für den tatsächlich vorliegenden Zustand zurückgeführt werden.

Basierend auf den Daten der untersuchten Vitalparameter entscheidet ein Alarmsystem online für jeden neuen Zeitpunkt, ob ein Alarm ausgelöst werden sollte. Bedingt durch Artefakte und Ausreißer kann es dabei zu Fehlalarmen kommen. Jede tatsächlich getroffene Entscheidung eines Alarmsystems kann retrospektiv mit Hilfe von Annotationen evaluiert werden. In den Annotationen ist festgehalten, welche Situationen alarmrelevant sind.

Jede zu bewertende Situation kann als eine Untersuchungseinheit aufgefasst werden, für die sowohl notiert wird, ob mit dem Alarmsystem ein Alarm gegeben wurde als auch, ob tatsächlich ein alarmrelevanter Zustand vorlag. Die Beobachtungen dieser zwei Merkmale können in Form einer Vierfelder-Tafel dargestellt werden. Für N Untersuchungseinheiten ergibt sich:

Tabelle 14.1

Alarmsystem	Zustand		Σ
	alarmrelevant	nicht alarmrelevant	
Alarm	N_{11} ↗	N_{12} ↘	N_1
kein Alarm	N_{21} ↘	N_{22} ↗	N_2
Σ	N_1	N_2	N

Ein ideales Alarmsystem besitzt eine hohe Richtig-Positiv (N_{11})- und Richtig-Negativ (N_{22})-Rate und geringe Falsch-Positiv (N_{12})- und Falsch-Negativ (N_{21})-Raten. Kenngrößen, die die Güte eines Diagnoseverfahrens beschreiben sind die Sensitivität und die Spezifität. Die Sensitivität gibt die Wahrscheinlichkeit dafür an, dass das Alarmsystem bei Vorliegen eines alarmrelevanten Zustands einen Alarm auslöst. Die Spezifität ist die Wahrscheinlichkeit dafür, dass das Alarmsystem im Falle eines nicht alarmrelevanten Zustands keinen Alarm

Version: 1.3	Letzte Änderung: 5.05.04	Seite 13 von 19
--------------	--------------------------	-----------------

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

auslöst. Die Schätzer für die Sensitivität und die Spezifität erhält man aus Tabelle 14.1 gemäß:

$$\widehat{\text{Sensitivität}} = N_{11} / (N_{11} + N_{21})$$

und

$$\widehat{\text{Spezifität}} = N_{22} / (N_{12} + N_{22})$$

Wird die Information aus den Beobachtungen mit Hilfe dieser Kenngrößen zusammengefasst, dann lautet die Fragestellung die durch das Hauptzielkriterium gegeben ist:

Besitzen die neu entwickelten Algorithmen im Intensivmonitoring eine höhere Spezifität bei einer unvermindert hohen Sensitivität gegenüber den in der Praxis eingesetzten Schwellwertalarmen?

Zu zeigen ist somit die „Nicht-Unterlegenheit“ des neu entwickelten Alarmsystems in Bezug auf die Sensitivität und die „Überlegenheit“ des neuen Systems in Bezug auf die Spezifität. Bezeichne im Folgenden A_1 das bisher in der Praxis eingesetzte Alarmsystem, das auf einfachen Schwellwertalarmen beruht, und A_2 das Alarmsystem, das auf neu entwickelten Algorithmen beruht.

Die Hypothesen für die beiden Kenngrößen Sensitivität und Spezifität lauten damit:

$H_{0\text{Sen}}$: Sensitivität(A_1) – Sensitivität(A_2) $\geq \delta$ vs.

$H_{1\text{Sen}}$: Sensitivität(A_1) – Sensitivität(A_2) $< \delta$, wobei $\delta = 0.05$ gewählt wird,

und

$H_{0\text{Spez}}$: Spezifität(A_1) – Spezifität (A_2) ≥ 0 vs. $H_{1\text{Spez}}$: Spezifität (A_1) – Spezifität (A_2) < 0 .

Da die zwei zu vergleichenden Algorithmen auf die gleichen Daten angewendet werden, handelt es sich um verbundene Stichproben. Eine Zusammenfassung der Beobachtungen in Form von zwei Kontingenztafeln wie in Tabelle 14.1 ist damit nur begrenzt aussagefähig (HAWASS, (1997) [27]). Geeigneter ist eine Darstellung gemäß der folgenden Tabelle 14.2, die zusätzlich aufzeigt, in wie vielen Fällen die Algorithmen die Situationen übereinstimmend bzw. unterschiedlich bewerten.

Tabelle 14.2

		Zustand				Σ	
		alarmrelevant		nicht alarmrelevant			
		$A_2 =$ Alarm	$A_2 =$ kein Alarm	$A_2 =$ Alarm	$A_2 =$ kein Alarm		
Zustand	alarm-relevant	$A_1 =$ Alarm	n_{111}	n_{110}	0	0	$n_{11.}$
		$A_1 =$ kein Alarm	n_{101}	n_{100}	0	0	$n_{10.}$
	nicht alarm-relevant	$A_1 =$ Alarm	0	0	n_{011}	n_{010}	$n_{01.}$
		$A_1 =$ kein Alarm	0	0	n_{001}	n_{000}	$n_{00.}$
		Σ	$n_{1.1}$	$n_{1.0}$	$n_{0.1}$	$n_{0.0}$	n

Tabelle 14.2 umfasst so genannte Symmetrietafeln, eine für die alarmrelevanten Zustände und eine für die nicht alarmrelevanten Zustände.

Aus diesem Schema ergeben sich die Schätzer für die Sensitivitäten und die Spezifitäten der

Version: 1.3	Letzte Änderung: 5.05.04	Seite 14 von 19
--------------	--------------------------	-----------------

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

beiden Verfahren gemäß:

$$\begin{aligned} \widehat{\text{Sensitivität}}(A_1) &= \frac{n_{111} + n_{110}}{n_{111} + n_{110} + n_{101} + n_{100}} & \text{und} & \quad \widehat{\text{Sensitivität}}(A_2) = \frac{n_{111} + n_{101}}{n_{111} + n_{110} + n_{101} + n_{100}} \\ \widehat{\text{Spezifität}}(A_1) &= \frac{n_{001} + n_{000}}{n_{011} + n_{010} + n_{001} + n_{000}} & \text{und} & \quad \widehat{\text{Spezifität}}(A_2) = \frac{n_{010} + n_{000}}{n_{011} + n_{010} + n_{001} + n_{000}} \end{aligned}$$

Gemäß der durch das Hauptzielkriterium gegebenen Fragestellung soll nun gezeigt werden, dass das Alarmsystem A_2 eine höhere Spezifität gegenüber Alarmsystem A_1 besitzt und gleichzeitig in Bezug auf die Sensitivität nicht unterlegen ist. Damit müssen Sensitivität und Spezifität simultan betrachtet werden, während in den obigen Hypothesen beide Kenngrößen einzeln getestet werden.

Um beide Hypothesen simultan zu prüfen formulieren LU, JIN und GENANT (2003) [42] das Testproblem wie folgt um:

$$H_0: H_{0\text{Sen}} \text{ oder } H_{0\text{Spez}} (= H_{0\text{Sen}} \cup H_{0\text{Spez}}) \text{ vs. } H_1: H_{1\text{Sen}} \text{ und } H_{1\text{Spez}} (= H_{1\text{Sen}} \cap H_{1\text{Spez}})$$

Damit wird gefordert, dass A_2 unter der Alternative sowohl eine nicht schlechtere Sensitivität als auch eine höhere Spezifität als A_1 aufweist.

Statistische Verfahren zur Auswertung

Zur Beantwortung der Fragestellung, die durch das Hauptzielkriterium gegeben ist, sind die Beobachtungen zunächst in Form von Vierfeldertafeln wie in Tabelle 14.1 bzw. Symmetrietafeln wie in Tabelle 14.2 zusammenzufassen. Die Untersuchungseinheiten bestehen im Rahmen dieser Studie aus kurzen Zeitabschnitten, für die jeweils festgehalten wird, ob in diesem Zeitraum von einem Alarmsystem ein Alarm ausgelöst wird und ob der Patientenzustand in diesem Zeitraum gemäß den Annotationen als alarmrelevant gilt. Die Länge der untersuchten Zeitsegmente ist noch festzulegen, wobei Abschnitte von m Minuten, mit m in etwa zwischen 2-10 Minuten, sinnvoll erscheinen.

Der vollständige Beobachtungszeitraum für einen Patienten k wird somit in n^k aufeinander folgende Bewertungszeiträume unterteilt. Der Gesamtstichprobenumfang n setzt sich schließlich aus der Summe sämtlicher Bewertungszeiträume für alle K Patienten zusammen, d.h. $n = n^1 + n^2 + \dots + n^K$.

Aus den so konstruierten Tabellen lassen sich Sensitivität und Spezifität wie gewohnt schätzen. Bei dem Vergleich mit ähnlichen Studien in der Literatur (siehe CHAMBRIN et al. (1999), LAWLESS (1994) und TSIEN (2000) [11, 40, 64]) ist jedoch Vorsicht geboten, da die Vierfeldertafeln für unterschiedliche Untersuchungseinheiten aufgestellt werden. Damit lassen sich i. a. die geschätzten Sensitivitäten und Spezifitäten aus verschiedenen Studien nicht direkt vergleichen.

Um für das Alarmsystem A_2 sowohl „Nicht-Unterlegenheit“ in Bezug auf die Sensitivität als auch „Überlegenheit“ in Bezug auf die Spezifität gegenüber dem Alarmsystem A_1 zu zeigen, ist das simultane Testproblem

$$H_0: H_{0\text{Sen}} \text{ oder } H_{0\text{Spez}} (= H_{0\text{Sen}} \cup H_{0\text{Spez}}) \text{ vs. } H_1: H_{1\text{Sen}} \text{ und } H_{1\text{Spez}} (= H_{1\text{Sen}} \cap H_{1\text{Spez}})$$

zu lösen. Da hier die Nullhypothese als Vereinigung und die Alternativhypothese als Schnitt formuliert ist, kann ein Intersection-Union-Test (IUT) konstruiert werden (BERGER (1996) [3]). Dieser Test führt genau dann zur Ablehnung der Nullhypothese H_0 zum Niveau α , wenn alle Hypothesen, aus denen sich die Nullhypothese zusammensetzt, einzeln zum Niveau α abgelehnt werden können. Zur Konstruktion des IUT müssen zunächst die Testprobleme $H_{0\text{Sen}}$

Version: 1.3	Letzte Änderung: 5.05.04	Seite 15 von 19
--------------	--------------------------	-----------------

SFB 475

Studienplan: Alarmstudie

vs. H_{1Sen} und H_{0Spez} vs. H_{1Spez} getrennt betrachtet werden.

Ein approximativer zweiseitiger Test auf Symmetrie für 2×2 -Tafeln ist der Test von McNemar (MCNEMAR (1947) [46]). Zur Überprüfung von einseitigen Hypothesen kann die zugehörige nicht quadrierte Teststatistik eingesetzt werden. Für das Nicht-Unterlegenheits-Testproblem H_{0Sen} vs. H_{1Sen} wird jedoch eine Verallgemeinerung der nicht quadrierten McNemar-Teststatistik, die sog. Score-Teststatistik von NAM (1997) [49] verwendet, welche gegeben ist durch:

$$S_1(\delta) = \frac{n_{110} - n_{101} - n_{1..} \delta}{\sqrt{n_{1..} (2\tilde{p}_{0||} + \delta - \delta^2)}},$$

wobei $n_{1..} = n_{111} + n_{110} + n_{101} + n_{100}$, $\tilde{p}_{0||} = \left(\sqrt{B^2 - 4AC} - B \right) / (2A)$, $A = 2n_{1..}$,
 $B = (2n_{1..} + n_{101} - n_{110})\delta - (n_{110} + n_{101})$ und $C = -n_{101}\delta(1 - \delta)$.

Unter H_{0Sen} ist $S_1(\delta)$ approximativ standardnormalverteilt. Die Hypothese H_{0Sen} wird somit zum Signifikanzniveau α abgelehnt, falls $S_1(\delta) < z_\alpha$, wobei z_α das α -Quantil der Standardnormalverteilung bezeichnet.

Zum Testen von H_{0Spez} vs. H_{1Spez} wird die einfache unquadrierte McNemar-Teststatistik

$$S_0 = \frac{n_{001} - n_{010}}{\sqrt{n_{001} + n_{010}}}$$

verwendet. Unter H_{0Spez} ist S_0 approximativ standardnormalverteilt. H_{0Spez} wird somit zum Signifikanzniveau α abgelehnt, wenn $S_0 < z_\alpha$.

Für den IUT zum Testen von H_0 vs. H_1 gilt nun, dass H_0 genau dann abgelehnt wird, wenn $S = \max(S_1(\delta), S_0) < z_\alpha$, da in diesem Fall die beiden in H_0 eingehenden Hypothesen abgelehnt werden können. Der IUT hat das globale Signifikanzniveau α , welches in dieser Studie auf $\alpha = 0.05$ festgelegt wird.

15 Zusammenfassung

Die geplante Studie dient der klinischen Validierung neuer Algorithmen zur Alarmgebung im Intensivmonitoring. Überprüft werden soll, ob mit neu entwickelten statistischen Online-Verfahren zur Alarmgebung die Fehlalarmrate bei unvermindert hoher Sensitivität gegenüber den in der Praxis eingesetzten Schwellwertalarmen reduziert werden kann. Zusätzlich sollen folgende Ziele verfolgt werden:

- Erstellung eines Referenzdatensatzes von Monitordaten und klinisch annotierten Alarmdaten.
- Untersuchung der Häufigkeit von Alarmen und der klinischen Relevanz.
- Überprüfung von neu entwickelten statistischen Verfahren zur Online-Mustererkennung.
- Untersuchung von Methoden der Dimensionsreduktion.

Hierzu sollen alle Daten von den Intensivmonitoren, d.h. numerische Messwerte, Echtzeitkurven, Alarmmeldungen und Alarmeinstellungen, über einen bestimmten Zeitraum gesammelt werden. Zusätzlich sollen bei mindestens 100 Patienten klinische Annotationen zur Bewertung der Alarmsituationen und des Patientenzustandes erhoben werden.

Eine Untersuchung von neuen Methoden direkt am Patienten oder mit unmittelbarem Einfluss auf die Diagnostik oder Therapie des Patienten ist nicht vorgesehen.

SFB 475

Studienplan: Alarmstudie

16 Literatur

1. Abelson, R. P., Tukey, J. W. (1963), "Efficient Utilization of Non-numerical Information in Quantitative Analysis: General Theory and the Case of Simple Order", *Annals of Mathematical Statistics*, 34, 1347-1369.
2. Bauer, M. (1997c), "Simultane Ausreißer- und Interventionsidentifikation in Online-Monitoring-Daten", Dissertation, Fachbereich Statistik, Universität Dortmund.
3. Berger RL, Hsu JC (1996): Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets. *Statistical Science* 11(4): 283-319
4. Berman JJ, Meyer J, Coats G (1984): Effects of program characteristics on treatment outcome: an interrupted time-series analysis. *J Stud Alcohol* 45: 405-410
5. Bernholt, T., Fried, R. (2003), "Computing the Update of the Repeated Median Regression Line in Linear Time", *Information Processing Letters* 88(3), 111-117.
6. Box GEP, Jenkins GM (1970): Time series analysis, forecasting and control. First edition. Holden-Day, San Francisco
7. Box GEP, Tiao GC (1975): Intervention analysis with applications to economic and environmental problems. *J Am Stat Assoc* 70: 70-79
8. Brillinger, D. R. (1989), "Consistent Detection of a Monotonic Trend Superposed by a Stationary Time Series", *Biometrika*, 76, 23-30
9. Brude U (1995): What is happening to the number of fatalities in road accidents? A model for forecasts and continuous monitoring of development up to the year 2000. *Accid Anal Prev* 27: 405-410
10. Brückner, A. (2001), "Robuste Schätzung der Autokovarianzfunktion", Diplomarbeit, Fachbereich Statistik, Universität Dortmund.
11. Chambrin MC, Ravoux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B (1999): Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med* 25: 1360-6
12. Cook TD, Campbell DT (1979): Quasi experimentation: Design and analysis issues for field settings. Rand McNally, Chicago
13. Cropp AJ, Woods LA, Raney D, Bredle DL (1994): Name that tone: the proliferation of alarms in the intensive care unit. *Chest* 105: 1217-20.
14. Davies, P. L., Fried, R., Gather, U. (2003), "Robust Signal Extraction for On-line Monitoring Data", *Journal of Statistical Planning and Inference* 122: 65-78
15. Fried, R. (2004), "Robust Filtering of Time Series with Trends". Proceedings of the Conference on Current Advances and Trends in Nonparametric Statistics, Special Issue of the "Journal of Nonparametric Statistics", erscheint.
16. Fried, R., Gather, U. (2002), "Fast and Robust Filtering of Time Series with Trends", *Proceedings in Computational Statistics COMPSTAT 2002*, Härdle, W., Rönz, B. (eds.), Physica, Heidelberg, 367-372.
17. Fried, R., Imhoff, M. (2004), "On the Online Detection of Monotonic Trends in Time Series". *Biometrical Journal* 46: 90-102
18. Gather, U., Bauer, M., Fried, R. (2002), "The Identification of Multiple Outliers in Online Monitoring Data", *Estatistica*, 54: 289-338.
19. Gather, U., Fried, R. (2003), "Robust Estimation of Scale for Local Linear Temporal Trends", *Tatra Mountains Mathematical Publications* 26: 87-101
20. Gerber WD, Schellenberg R, Thom M, et al (1995): Cyclandelate versus propranolol in the prophylaxis of migraine--a double-blind placebo-controlled study. *Funct Neurol* 10: 27-35
21. Gordon K (1986): The multi-state Kalman Filter in medical monitoring. *Comput Methods Programs Biomed* 23: 147-154

Version: 1.3

Letzte Änderung: 5.05.04

Seite 17 von 19

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

22. Gordon K, Smith AFM (1990): Modeling and monitoring biomedical time series. *J Am Stat Ass* 85: 328-337
23. Grübel, R. (1988), "The Length of the Shorth", *Annals of Statistics*, 16, 619-628.
24. Haimowitz I, Kohane IS (1996): Managing temporal worlds for medical trend diagnosis. *Art Intel Med* 8: 299-321
25. Haimowitz I, Le PP, Kohane IS (1995): Clinical monitoring using regression-based trend templates. *Art Intel Med* 7: 473-496
26. Hampel, F. R., (1975), "Beyond Location Parameters: Robust Concepts and Methods", *Bulletin of the Int. Statist. Inst.*, 46, 375-382.
27. Hawass NED (1997): Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients. *The British Journal of Radiology* 70: 360-366
28. Helfenstein U (1987): Zeitreihenanalyse eines Praeventionseffektes am Beispiel des Unfallgeschehens in Zürich vor und nach Tempo 50. *Soz Praventivmed* 32: 305-309
29. Hepworth JT, Hendrickson SG, Lopez J (1994) Time series analysis of physiological response during ICU visitation. *West J Nurs Res* 16: 704-717
30. Hill DW, Endresen J (1978): Trend recording and forecasting in intensive care therapy. *Br J Clin Equipment* 1978: 5-14
31. Holm K (1983): Single subject research. *Nurs Res* 32: 253-255
32. Imhoff M, Bauer M (1996): Time series analysis in critical care monitoring. *New Horiz* 4: 519-531
33. Imhoff M, Bauer M, Gather U, Löhlein D (1997) Time series analysis in intensive care medicine. *Appl Cardiopulmonary Pathophysiol* 6: 263-281
34. Imhoff M, Bauer M, Gather U, Löhlein D (1998): Statistical pattern detection in univariate time series of intensive care on-line monitoring data. *Intensive Care Med* 24: 1305-1314
35. Imhoff, M., Bauer, M., Gather, U., Fried, R. (2002), "Pattern Detection in Intensive Care Monitoring Time Series with Autoregressive Models: Influence of the Model Order", *Biometrical Journal*, 44, 746-761.
36. Imhoff M, Lehner J, Löhlein D (1989): Time series analysis in the postoperative follow-up after liver resections. 10th International Meeting on Clinical Biostatistics. The International Society for Clinical Biostatistics, Maastricht, 11.-14.9.1989
37. Koski EM, Makivirta A, Sukuvaara T, Kari A (1990): Frequency and reliability of alarms in the monitoring of cardiac postoperative patients. *Int J Clin Monit Comput* 7: 129-33
38. Koski EM, Makivirta A, Sukuvaara T, Kari A (1995): Clinicians' opinions on alarm limits and urgency of therapeutic responses. *Int J Clin Monit Comput* 12: 85-8.
39. Lambert CR, Raymenants E, Pepine CJ (1995): Time-series analysis of long-term ambulatory myocardial ischemia: effects of beta-adrenergic and calcium channel blockade. *Am Heart J* 129: 677-684
40. Lawless, S.T. (1994): Crying wolf: False alarms in a pediatric intensive care unit. *Critical Care Med* 22:981-985
41. Lim KL, Muir K, Powell RJ (1994): Urine neopterin: a new parameter for serial monitoring of disease activity in patients with systemic lupus erythematosus. *Ann Rheum Dis* 53: 743-748
42. Lu Y, Jin H, Genant HK (2003): On the non-inferiority of a diagnostic test based on paired observations. *Statistics in Medicine* 22, 3029-3044
43. Ludwig, M. (2001), "Online Trenderkennung in Zeitreihen", Diplomarbeit, Fachbereich Statistik, Universität Dortmund.
44. Mäkivirta, A., Koski, E., Kari, A. and Sukuvaara, T. (1991): The median filter as a preprocessor for a patient monitor limit alarm system in intensive care. *Comput Methods Programs Biomed* 34: 139-144

Version: 1.3	Letzte Änderung: 5.05.04	Seite 18 von 19
--------------	--------------------------	-----------------

SFB 475	Studienplan: Alarmstudie
---------	--------------------------

45. Marrazzi MA, Markham KM, Kinzie J, et al (1995): Binge eating disorder: response to naltrexone. *Int J Obes Relat Metab Disord* 19: 143-145
46. McNemar Q (1947): Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12: 153-157
47. Meredith C, Edworthy J (1995): Are there too many alarms in the intensive care unit? An overview of the problems. *J Adv Nurs* 21: 15-20
48. Morris AH (1999): Computerized protocols and bedside decision support. *Crit Care Clin* 15: 523-545
49. Nam JM (1997): Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics* 53: 1422-1430
50. O'Carroll, T.M. (1986): Survey of Alarms in an Intensive Therapy Unit. *Anaesthesia* 41 , 742-744
51. Rock SM (1995): Impact of the 65 mph speed limit on accidents, deaths, and injuries in Illinois. *Accid Anal Prev* 27: 207-214
52. Rosel J, Elosegui E (1994): Daily and weekly smoking habits: a Box-Jenkins analysis. *Psychol Rep* 75: 1639-1648
53. Rousseeuw, P. J. (1984), "Least Median of Squares Regression", *Journal of the American Statistical Association*, 79, 871-880.
54. Rousseeuw, P. J., Croux, C. (1993), "Alternatives to the Median Absolute Deviation", *Journal of the American Statistical Association*, 88, 1273-1283.
55. Rousseeuw, P. J., Leroy, A. M. (1988), "A Robust Scale Estimator Based on the Shortest Half", *Statistica Neerlandica*, 42, 103-116.
56. Rotton J, Frey J (1985): Air pollution, weather, and violent crimes: concomitant time-series analysis of archival data. *J Pers Soc Psychol* 49: 1207-1220
57. Saez M, Sunyer J, Castellsague J, et al (1995): Relationship between weather temperature and mortality: a time series analysis approach in Barcelona. *Int J Epidemiol* 24: 576-582
58. Shewhart WA (1931): *Economic Control of Quality Manufactured Product*. D. Van Nostrand Reinhold, Princeton NJ
59. Siegel, A. F. (1982), "Robust Regression Using Repeated Medians", *Biometrika*, 68, 242-244.
60. Smith AFM, West M (1983): Monitoring renal transplants: an application of the multi-process Kalman filter. *Biometrics* 39: 867-878
61. Smith AFM, West M, Gordon K, et al (1983): Monitoring kidney transplant patients. *Statistician* 32: 46-54
62. Touloumi G, Pocock SJ, Katsouyanni K (1994): Short-term effects of air pollution on daily mortality in Athens: a time-series analysis. *Int J Epidemiol* 23: 957-967
63. Trigg DW (1964): Monitoring of a forecasting system. *Operational Research Quarterly* 15: 271-274
64. Tsien MC (2000): *TrendFinder: Automated detection of alarmable trends*. PhD thesis, Massachusetts Institute of Technology
65. Vingilis E, Blefgen H, Lei H, et al (1988): An evaluation of the deterrent impact of Ontario's 12-hour licence suspension law. *Accid Anal Prev* 20: 9-17
66. Wagenaar AC, Langley JD (1995): Alcohol licensing system changes and alcohol consumption: introduction of wine into New Zealand grocery stores. *Addiction* 90: 773-783
67. Wagenaar AC, Webster DW (1986): Preventing injuries to children through compulsory automobile safety seat use. *Pediatrics* 78: 662-672

Version: 1.3	Letzte Änderung: 5.05.04	Seite 19 von 19
--------------	--------------------------	-----------------

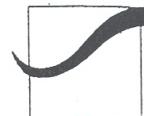


Zentrum für Klinische Studien
Klinikum der Universität Regensburg D-93042 Regensburg

Herrn
Dr. Reng
Klinik und Poliklinik für Innere Medizin I

im Hause

UNIVERSITÄT
REGENSBURG



KLINIKUM

Zentrum für Klinische Studien

**Geschäftsstelle der Ethik-Kommission der
Medizinischen Fakultät der Universität Regensburg**

Vorsitzender:

Prof. Dr. med. R. Andreesen

Telefon: +49 - (0)941 - 944-5631

Telefax: +49 - (0)941 - 944-5632

ethikkommission@klinik.uni-regensburg.de

15. Okt. 04

Nr. 04/094

Validierung von Alarmalgorithmen in der Intensivmedizin

Amendment 1 vom 07.10.2004 incl. Patienten/Verwandten-Information.

Sehr geehrter Herr Dr. Reng,

die Ethik-Kommission nimmt das mit Schreiben vom 07.10.2004 übersandte Amendment zu o. g. Studie zur Kenntnis. An der positiven Bewertung der Studie ändert sich nichts.

Mit freundlichen Grüßen

Prof. Dr. R. Andreesen

Hausanschrift:
Zentrum für Klinische Studien
Klinikum der Universität Regensburg
Franz-Josef-Strauß-Allee 11
D-93053 Regensburg

Bankverbindung
Sparkasse Regensburg
Konto Nr. 780 010 500
BLZ 750 500 00

ethikkommission@klinik.uni-regensburg.de
<http://www.zks-regensburg.de>

Amendment 1 zur Studie

„Validierung von Alarmalgorithmen in der Intensivmedizin“

Im Versuchsplan ist unter Punkt [9] die Sammlung der klinischen Daten durch einen Beobachter vorgesehen.

„Der Beobachter wird jeweils 8 Stunden kontinuierlich einen oder zwei Patienten (in einem Zwei-Bett-Zimmer) beobachten.“

„Der klinische Beobachter erfasst mit Hilfe des Studiennotebooks und einer entsprechenden Eingabemaske folgende Beobachtungen [...]“

Die online-Dokumentation der Alarme mit Erfassung der klinischen Situation durch einen Studienbeobachter zeigt sich in der Praxis aufgrund fehlender personeller Kapazitäten als nicht umsetzbar. Eine lückenlose Erfassung der klinischen Situation ist aber Grundvoraussetzung für die Durchführung dieser Untersuchung.

Daher soll die Änderung des Versuchsplanes wie folgt angepasst werden:

„Die Erfassung der klinischen Situation erfolgt mit Hilfe einer im Patientenzimmer angebrachten Kamera. Die hierdurch erfassten klinischen Daten werden zeitnah durch einen Studienbeobachter ausgewertet und mit den dokumentierten Alarmen korreliert. Nach der Dokumentation erfolgt die Vernichtung der durch die Kamera aufgezeichneten Daten.“

Durch die hiermit mögliche personell unabhängige Dokumentation, die eine methodische Verbesserung der Studie bedeutet, und um einen möglichen Datenverlust durch z.B. technische Mängel möglichst gering zu halten, soll die gefilmte Beobachtungsperiode pro Patient auf 24 Stunden anstelle 8 Stunden ausgeweitet werden.

Somit soll folgender Satz in den Versuchsplan eingefügt werden:

„Die Videoaufzeichnung der klinischen Daten erfolgt für jeweils 24 Stunden.“

Die Information der Patienten und ihrer Verwandten über die Studie und die zwischenzeitliche Erfassung persönlicher Daten mittels Kamera erfolgt durch ein entsprechendes Informationsblatt.

Liebe Patienten und Angehörige der Station 92

Unsere Aufgabe ist die bestmögliche Versorgung unserer Patienten. Um diese Versorgung noch weiter zu verbessern, ist es für uns als Universitätsklinik wichtig, Studien durchzuführen. Nur so ist die Optimierung und ständige Weiterentwicklung der Patientenversorgung möglich.

Auf unserer Intensivstation findet zur Zeit eine Studie statt, die sich mit dem Auftreten von Alarmen an den Bettplatzmonitoren beschäftigt. Ziel dieser Studie ist es, die Häufigkeit der Alarme des Monitors über dem Bett zu verringern, ohne dabei wichtige Alarme zu übersehen. Hierzu ist es jedoch notwendig, die Situation, unter der es zu einem Alarm gekommen ist, genau und lückenlos zu analysieren.

Zu diesem Zweck wurde in einem Zimmer der Intensivstation Kameras angebracht, durch die der Bettplatz und der Monitor aufgenommen werden können. Die Aufzeichnungen werden benötigt, um die Alarme richtig auswerten zu können. Nach Beendigung der Aufzeichnung werden die Filme von einem Arzt dieser Station angesehen, bewertet und danach wieder vernichtet.

Unser Interesse liegt bei der Auswertung der Informationen auf der Beurteilung der verschiedenen Alarme, die im Laufe eines Tages auftreten, und die sich daraus ergebende Reaktion unsererseits. Persönliche Daten bzw. Informationen über Sie als Patient oder über Ihren Angehörigen werden nicht erfasst und finden keinen Eingang in die Studie. Auch wird die Behandlung unserer Patienten bzw. Ihres Angehörigen hierdurch in keiner Weise beeinflusst.

Wir möchten uns schon jetzt bei Ihnen ganz herzlich für Ihr Verständnis bedanken.

Dr. med. Julia Langgartner
Oberärztin der Station 92

Dr. med. Christian Wrede
Oberarzt der Station 92

Anhang B

Ergänzende Deskription des Referenzdatensatzes

Fall	Morgen (6-12 Uhr)				Nachmittag (12-18 Uhr)				Abend (18-0 Uhr)				Nacht (0-6 Uhr)			
	Dauer	Ant.	A	S	Dauer	Ant.	A	S	Dauer	Ant.	A	S	Dauer	Ant.	A	S
1	-	-	-	-	03:00:01	1.00	11	8	-	-	-	-	-	-	-	-
2	02:45:01	0.26	6	6	-	-	-	-	02:00:00	0.19	2	1	05:59:52	0.56	6	4
3	-	-	-	-	02:30:29	0.17	7	2	06:00:00	0.41	1	0	05:57:45	0.41	4	0
4	04:53:10	0.24	6	2	06:00:00	0.30	2	1	05:59:35	0.30	3	0	03:21:59	0.17	14	0
5	00:55:01	1.00	8	2	-	-	-	-	-	-	-	-	-	-	-	-
6	05:03:36	0.55	18	8	04:09:31	0.45	7	6	-	-	-	-	-	-	-	-
7	06:11:33	0.26	2	0	06:00:00	0.25	2	1	05:59:10	0.25	1	0	06:00:00	0.25	1	0
8	02:11:00	0.11	3	1	05:45:10	0.29	2	1	06:00:00	0.30	3	2	06:00:00	0.30	2	2
9	-	-	-	-	05:27:36	0.38	28	23	06:00:00	0.41	26	23	03:06:01	0.21	35	31
10	04:50:56	0.34	29	21	06:00:00	0.42	21	16	03:30:01	0.24	5	3	-	-	-	-
11	04:54:21	0.61	27	21	03:10:01	0.39	32	26	-	-	-	-	-	-	-	-
12	03:52:03	0.41	2	0	05:29:27	0.59	8	2	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	02:52:05	1.00	2	1	-	-	-	-
14	04:24:53	0.32	10	6	06:00:00	0.43	16	8	03:23:01	0.25	4	4	-	-	-	-
15	04:20:11	0.49	5	4	04:34:41	0.51	4	3	-	-	-	-	-	-	-	-
16	04:32:27	0.22	8	7	06:00:00	0.29	4	3	06:00:00	0.29	22	21	03:55:01	0.19	28	27
17	02:47:55	0.54	12	6	02:25:05	0.46	2	2	-	-	-	-	-	-	-	-
18	01:00:01	0.06	10	4	04:51:55	0.27	5	4	06:00:00	0.34	3	1	06:00:00	0.34	1	1
19	04:49:22	0.70	10	6	02:03:50	0.30	10	3	-	-	-	-	-	-	-	-
20	03:14:32	0.78	9	9	00:55:01	0.22	3	3	-	-	-	-	-	-	-	-
21	03:16:59	0.60	6	6	02:12:31	0.40	10	9	-	-	-	-	-	-	-	-
22	03:17:42	0.56	4	1	02:37:46	0.44	9	5	-	-	-	-	-	-	-	-
23	05:04:42	0.45	7	5	06:00:00	0.53	1	1	00:09:04	0.01	0	0	-	-	-	-
24	04:18:10	0.19	3	1	06:00:00	0.27	2	0	06:00:00	0.27	1	0	06:00:00	0.27	4	2
25	06:00:00	0.25	2	0	06:00:00	0.25	4	4	06:00:00	0.25	2	0	06:00:00	0.25	2	0
26	01:33:13	0.09	1	1	02:58:31	0.18	4	1	06:00:00	0.36	3	1	06:00:00	0.36	8	2
27	06:42:01	0.27	10	9	06:00:00	0.24	3	3	06:00:00	0.24	1	1	06:00:00	0.24	2	2
28	06:41:31	0.28	10	8	05:15:03	0.22	4	4	06:00:00	0.25	4	3	06:00:00	0.25	2	2
29	06:56:54	0.28	12	8	06:00:00	0.24	5	4	06:00:00	0.24	3	2	06:00:00	0.24	5	3
30	04:14:40	0.36	7	4	06:00:00	0.51	12	9	01:30:01	0.13	9	7	-	-	-	-
31	04:42:54	0.21	2	1	06:00:00	0.26	7	4	06:00:00	0.26	12	8	06:00:00	0.26	9	7
32	06:00:00	0.26	4	2	04:51:51	0.21	13	11	06:00:00	0.26	9	8	06:00:00	0.26	9	7
33	02:43:19	0.43	1	0	03:35:01	0.57	7	2	-	-	-	-	-	-	-	-
34	04:16:20	0.53	15	10	03:51:01	0.47	17	12	-	-	-	-	-	-	-	-
35	02:38:58	0.51	3	2	02:30:57	0.49	0	0	-	-	-	-	-	-	-	-
36	00:49:40	0.05	4	2	02:52:15	0.18	8	5	06:00:00	0.38	2	1	06:00:00	0.38	8	6
37	01:08:08	0.08	3	2	01:50:15	0.12	13	8	06:00:00	0.40	5	4	06:00:00	0.40	1	1
38	06:00:00	0.37	6	3	00:36:59	0.04	0	0	03:45:00	0.23	0	0	06:00:00	0.37	1	0
39	01:28:02	0.08	3	1	03:50:05	0.22	2	2	06:00:00	0.35	1	0	06:00:00	0.35	2	0
40	04:13:45	0.19	5	1	06:00:00	0.27	5	5	06:00:00	0.27	6	1	06:00:00	0.27	2	0
41	01:41:55	0.10	2	1	03:42:21	0.21	1	0	06:00:00	0.34	1	0	06:00:00	0.34	0	0
42	01:23:53	0.07	0	0	06:00:00	0.31	2	2	06:00:00	0.31	1	0	06:00:00	0.31	1	0
43	06:11:03	0.26	2	1	06:00:00	0.25	3	1	06:00:00	0.25	0	0	06:00:00	0.25	0	0
44	05:58:12	0.26	10	5	06:00:00	0.26	4	3	04:59:46	0.22	4	4	06:00:00	0.26	2	1
45	04:44:29	0.44	5	3	06:00:00	0.55	2	1	00:04:57	0.01	0	0	-	-	-	-
46	05:52:24	0.23	4	4	06:00:00	0.24	10	8	08:52:47	0.35	2	2	04:30:37	0.18	2	2
47	05:52:21	0.23	2	1	06:00:00	0.24	1	1	08:53:04	0.35	1	0	04:31:50	0.18	1	1
48	05:00:01	0.28	5	2	00:34:57	0.03	14	5	06:00:00	0.34	2	2	06:00:00	0.34	6	4
49	-	-	-	-	01:07:55	0.89	1	1	00:08:45	0.11	0	0	-	-	-	-
50	-	-	-	-	03:09:50	0.35	4	3	06:00:00	0.65	2	1	-	-	-	-
51	-	-	-	-	03:10:27	0.27	1	1	06:00:00	0.51	5	4	02:30:01	0.21	4	4
52	06:00:00	0.25	2	0	06:00:00	0.25	1	0	06:00:00	0.25	1	0	06:00:00	0.25	1	0
53	06:00:00	0.83	8	5	01:15:53	0.17	3	0	-	-	-	-	-	-	-	-
54	00:32:33	0.41	4	0	00:46:28	0.59	8	1	-	-	-	-	-	-	-	-
55	06:00:00	0.32	16	12	06:00:00	0.32	8	6	05:57:34	0.32	4	3	00:40:31	0.04	4	3
56	04:04:01	0.57	6	5	-	-	-	-	-	-	-	-	03:06:40	0.43	1	0
57	06:00:00	0.25	2	1	06:00:00	0.25	14	10	06:00:00	0.25	7	6	05:38:21	0.24	1	1
58	03:41:50	0.24	4	2	-	-	-	-	05:53:55	0.38	2	2	06:00:00	0.38	2	2
59	02:16:16	0.38	4	3	-	-	-	-	-	-	-	-	03:44:32	0.62	2	1
60	-	-	-	-	-	-	-	-	03:08:49	0.39	5	4	05:00:01	0.61	3	2
61	-	-	-	-	-	-	-	-	01:58:42	1.00	11	5	-	-	-	-
62	01:45:00	0.09	4	1	06:00:00	0.30	8	5	06:00:00	0.30	14	11	06:00:00	0.30	4	3
63	03:43:31	0.17	2	1	06:00:00	0.28	7	6	06:00:00	0.28	4	3	06:00:00	0.28	9	7
64	06:00:00	0.26	42	41	06:00:00	0.26	13	11	05:24:15	0.23	21	13	06:00:00	0.26	7	5
65	05:08:10	0.36	16	13	06:00:00	0.42	8	7	03:08:31	0.22	2	0	-	-	-	-
66	00:16:38	0.03	7	0	-	-	-	-	02:51:21	0.31	13	12	06:00:00	0.66	4	3
67	06:00:00	0.31	1	0	06:00:00	0.31	0	0	03:52:13	0.20	1	1	03:13:51	0.17	1	1
68	06:00:00	0.25	3	2	06:00:00	0.25	3	1	06:00:00	0.25	4	4	06:14:32	0.26	5	3
Insg.	243:05:17	0.25	8	6	263:12:53	0.27	7	5	252:22:36	0.26	5	4	223:31:34	0.23	4	3

Tabelle B.1: Beobachtungsdauer (Dauer) und Anteil (Ant.) für die verschiedenen Tageszeiten sowie Alarme pro Stunde (A) und Schwellwertalarme pro Stunde (S)

Parameter	Anzahl Fälle	Parameter	Anzahl Fälle	Parameter	Anzahl Fälle
HR	68	NBP.S	8	PA.D	4
PLS	68	CVP	7	PA.M	4
SpO2	68	STaVR	6	RRc	3
ART.M	68	ICP	6	etCO2	3
ART.S	68	CPP	6	LA	2
ART.D	68	PIP	6	iCO2	2
ARR	68	PEEP	6	PWP	2
PVC.min	65	MAP	6	C.O	2
Ta	55	MVe	6	IT	2
RESP	37	TVe	6	BT	2
STaVL	37	RRv	6	C.O.1	2
STV	35	Pause	6	SVR	1
STIII	32	iO2	6	SvO2	1
STII	13	MVi	6	BT	1
STI	9	TVi	6	CCO	1
STaVF	9	InspT	6		
NBP.M	8	X.Pace	5	ART (2)	10
NBP.D	8	PA.S	4	Weitere	15

Tabelle B.2: Anzahl Fälle, bei denen der entsprechende Vitalparameter vom Monitoring-Gerät gemessen wurde. ART (2) bedeutet, dass eine (oder mehrere) zusätzliche arterielle Blutdruckmessung(en) an einer anderen Stelle durchgeführt wurde(n). Unter „Weitere“ wurden alle nicht näher bezeichneten Variablen zusammengefasst.

Fall	ARR	ART.M	HR	SpO2	ART.S	PVC/min	ART.M (2)	NBP.M	NBP.S	RESP	ART.S (2)	PA.M
1	X	X	X	X	X	X						
2	X	X	X	X	X	X						
3	X	X	X	X	X	X						
4	X	X	X	X	X	X						
5	X	X	X	X	X	X						
6	X	X	X	X	X	X						X
7	X	X	X	X	X	X		X	X			
8	X	X	X	X	X	X						
9	X	X	X	X	X	X				X		
10	X	X	X	X	X	X						
11	X	X	X	X	X	X						
12	X	X	X	X	X	X						
13	X	X	X	X	X	X						
14	X	X	X	X	X	X				X		
15	X	X	X	X	X	X						
16	X	X	X	X	X	X				X		
17	X	X	X	X	X	X		X	X			
18	X	X	X	X	X	X						
19	X	X	X	X	X	X						
20	X	X	X	X	X	X						
21	X	X	X	X	X	X		X	X			
22	X	X	X	X	X	X						
23	X	X	X	X	X	X						
24	X	X	X	X	X	X						
25	X	X	X	X	X	X						
26	X	X	X	X	X	X	X				X	
27	X	X	X	X	X	X	X					
28	X	X	X	X	X	X				X		
29	X	X	X	X	X	X	X					
30	X	X	X	X	X	X						X
31	X	X	X	X	X	X	X				X	
32	X	X	X	X	X	X	X				X	
33	X	X	X	X	X	X						
34	X	X	X	X	X	X				X		
35	X	X	X	X	X	X						
36	X	X	X	X	X	X	X					
37	X	X	X	X	X	X						
38	X	X	X	X	X	X	X				X	X
39	X	X	X	X	X	X						
40	X	X	X	X	X	X						
41	X	X	X	X	X	X						
42	X	X	X	X	X	X						
43	X	X	X	X	X							
44	X	X	X	X	X	X		X	X			
45	X	X	X	X	X	X						
46	X	X	X	X	X	X						
47	X	X	X	X	X	X						
48	X	X	X	X	X	X						
49	X	X	X	X	X	X						
50	X	X	X	X	X	X	X				X	
51	X	X	X	X	X	X						
52	X	X	X	X	X	X						
53	X	X	X	X	X	X				X		
54	X	X	X	X	X	X		X	X			
55	X	X	X	X	X	X				X		
56	X	X	X	X	X	X		X	X			
57	X	X	X	X	X	X		X	X			
58	X	X	X	X	X	X						
59	X	X	X	X	X	X						
60	X	X	X	X	X	X						
61	X	X	X	X	X	X		X	X			
62	X	X	X	X	X	X				X		
63	X	X	X	X	X	X						
64	X	X	X	X	X	X	X				X	
65	X	X	X	X	X	X						
66	X	X	X	X	X	X						X
67	X	X	X	X	X	X						
68	X	X	X	X	X	X						
Insgesamt	68	68	68	68	67	63	9	8	8	8	6	4

Tabelle B.3: Überwachte Variablen für die Fälle im Referenzdatensatz

Vitalparameter	Meldung	Anzahl	Anteil
Inv. Druck	Werte unterhalb des Messbereichs	536	0.6140
Inv. Druck	Statischer Druck	232	0.2658
Inv. Druck	Werte oberhalb des Messbereichs	74	0.0848
RESP	Koinzidenz	31	0.0355
Insgesamt		873	1.0000

Tabelle B.4: Alarmursachen für sonstige ernste Alarme

Vitalparameter	Meldung	Anzahl	Anteil
SpO2	Artefakt	250	0.3098
SpO2	Transparent	189	0.2342
SpO2	Schwaches Signal	154	0.1908
SpO2	Keine Messung	112	0.1388
HR	EKG-Ableitungen ungültig	50	0.0620
SpO2	Abgezogen	18	0.0223
HR	EKG Artefakt	7	0.0087
SpO2	Kein Lichtdurchlass	5	0.0062
NBP	Schlauch abgezogen	5	0.0062
SpO2	Sensor nicht kompatibel	4	0.0050
Inv. Druck	Hemo-Modul nicht angeschlossen	4	0.0050
NBP	Keine Pulsation	2	0.0025
RESP	Artefakt	2	0.0025
Inv. Druck	Pod 2 nicht angeschlossen	2	0.0025
Inv. Druck	Pod 3 nicht angeschlossen	1	0.0012
Inv. Druck	Hardwarefehler	1	0.0012
SpO2	Regulierungsfehler	1	0.0012
Insgesamt		807	1.0000

Tabelle B.5: Alarmursachen für technische Alarme

Vitalparameter	Meldung	Anzahl	Anteil
ARR	Bradykardie	66	0.6111
ARR	Supraventrikuläre Tachykardie	19	0.1759
ARR	Ventrikulärer Run	13	0.1204
HR	Asystolie	5	0.0463
ARR	Beschleunigter ventrikulärer Rhythmus	4	0.0370
ARR	Asystolie	1	0.0093
Insgesamt		108	1.0000

Tabelle B.6: Alarmursachen für Arrhythmie-Alarme

Vitalparameter	Meldung	Anzahl	Anteil
ART.S	Untere Schwellwertverletzung	941	0.1586
ART.S	Obere Schwellwertverletzung	916	0.1544
SpO2	Untere Schwellwertverletzung	784	0.1321
HR	Obere Schwellwertverletzung	560	0.0944
ART.S	Werte unterhalb des Messbereichs	495	0.0834
ART.M	Untere Schwellwertverletzung	350	0.0590
SpO2	Artefakt	250	0.0421
ART.S	Statischer Druck	198	0.0334
SpO2	Transparent	189	0.0319
HR	Untere Schwellwertverletzung	167	0.0281
SpO2	Schwaches Signal	154	0.0260
RESP	Obere Schwellwertverletzung	153	0.0258
ART.M	Obere Schwellwertverletzung	135	0.0228
SpO2	Keine Messung	112	0.0189
ARR	Bradykardie	66	0.0111
HR	EKG-Ableitungen ungültig	50	0.0084
ART.S	Werte oberhalb des Messbereichs	50	0.0084
PA.M	Untere Schwellwertverletzung	47	0.0079
RESP	Koinzidenz	31	0.0052
ART.S (2)	Untere Schwellwertverletzung	28	0.0047
PA.M	Werte oberhalb des Messbereichs	24	0.0040
ARR	Supraventrikuläre Tachykardie	19	0.0032
SpO2	Abgezogen	18	0.0030
ART.M (2)	Obere Schwellwertverletzung	17	0.0029
ART.M (2)	Statischer Druck	17	0.0029
ART.S (2)	Werte unterhalb des Messbereichs	14	0.0024
ARR	Ventrikulärer Run	13	0.0022
PVC/min	Obere Schwellwertverletzung	12	0.0020
ART.M	Werte unterhalb des Messbereichs	11	0.0019
ART.M (2)	Werte unterhalb des Messbereichs	10	0.0017
RESP	Untere Schwellwertverletzung	10	0.0017
HR	EKG-Artefakt	7	0.0012
NBP.S	Untere Schwellwertverletzung	7	0.0012
ART.M	Statischer Druck	7	0.0012
ART.M (2)	Untere Schwellwertverletzung	7	0.0012
PA.M	Obere Schwellwertverletzung	6	0.0010
PA.M	Statischer Druck	6	0.0010
PA.M	Werte unterhalb des Messbereichs	6	0.0010
ART.S (2)	Obere Schwellwertverletzung	5	0.0008
HR	Asystolie	5	0.0008

Fortsetzung auf der nächsten Seite

Tabelle B.7: Häufigkeiten der ausgelösten Alarmmeldungen

Vitalparameter	Meldung	Anzahl	Anteil
NBP.S	Schlauch abgezogen	5	0.0008
SpO2	Kein Lichtdurchlass	5	0.0008
ART.S (2)	Statischer Druck	4	0.0007
ARR	Beschleunigter ventrikulärer Rhythmus	4	0.0007
SpO2	Sensor nicht kompatibel	4	0.0007
PA.D	Hemo-Modul nicht angeschlossen	3	0.0005
ART.S	Pod 2 nicht angeschlossen	2	0.0003
NBP.S	Keine Pulsation	2	0.0003
RESP	Artefakt	2	0.0003
ARR	Asystolie	1	0.0002
ART.S	Hardwarefehler	1	0.0002
ART.S	Hemo-Modul nicht angeschlossen	1	0.0002
ART.S	Pod 3 nicht angeschlossen	1	0.0002
NBP.M	Obere Schwellwertverletzung	1	0.0002
SpO2	Regulierungsfehler	1	0.0002
Insgesamt		5934	1.0000

Tabelle B.7: Häufigkeiten der ausgelösten Alarmmeldungen

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe	
		ja	nein	hinw.		
richtig	nein	398	285	1912	2595	
		0.15	0.11	0.74	1.00	
		0.90	0.38	0.86	0.76	
		0.07	0.05	0.32	0.44	
	ja	42	474	311	827	
		0.05	0.57	0.38	1.00	
		0.10	0.62	0.14	0.24	
		0.01	0.08	0.05	0.14	
		Summe	440	759	2223	3422
			0.13	0.22	0.65	1.00
			0.50	0.28	1.00	0.59
			0.07	0.13	0.37	0.58
	falsch	nein	414	346	-	760
0.54			0.46	-	1.00	
0.93			0.18	-	0.32	
0.07			0.06	-	0.13	
ja		31	1607	-	1638	
		0.02	0.98	-	1.00	
		0.07	0.82	-	0.68	
		0.01	0.27	-	0.28	
		Summe	445	1953	-	2398
			0.19	0.81	-	1.00
			0.50	0.72	-	0.41
			0.07	0.33	-	0.40
Insgesamt		nein	812	631	1912	3355
	0.24		0.19	0.57	1.00	
	0.92		0.23	0.86	0.58	
	0.14		0.11	0.32	0.57	
	ja	73	2081	311	2465	
		0.03	0.84	0.13	1.00	
		0.08	0.77	0.14	0.42	
		0.01	0.35	0.05	0.42	
		Summe	885	2712	2223	5820
			0.15	0.47	0.38	1.00
			1.00	1.00	1.00	1.00
			0.15	0.46	0.37	0.98
					Nicht zu beurteilen	55
				Nicht bewertet	59	
				Alarme insgesamt	5934	

Tabelle B.8: Alarm-Annotationen der ausgelösten Alarme. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Ursache	Anzahl	Anteil an technisch falschen Alarmen
Sonstige Signalprobleme	1447	0.6034
Lagerung	515	0.2148
Bewegung(en) Patient	263	0.1097
Blutabnahme	171	0.0713
Nullabgleich	71	0.0296
Dislokation	57	0.0238
Leitung abgeknickt	54	0.0225
EKG zählt falsch	43	0.0179
Elektrode ab	7	0.0029
Fehlerhafte Platzierung	3	0.0013
Niedervoltage	1	0.0004
Perfusionsstörung	1	0.0004

Tabelle B.9: Ursachen für technisch falsche Alarme. Bei einem Alarm können mehrere Ursachen gleichzeitig auftreten.

Ursache	EKG		HR		Inv. Druck		NIBP		SpO2	
	Anz.	Ant.	Anz.	Ant.	Anz.	Ant.	Anz.	Ant.	Anz.	Ant.
Sonst. Signalprobleme	34	0.172	19	0.328	1116	0.711	7	0.875	271	0.480
Lagerung	20	0.101	11	0.190	336	0.214	2	0.250	146	0.258
Bewegung(en) Patient	5	0.025	1	0.017	52	0.033	4	0.500	201	0.356
Blutabnahme	0	0.000	0	0.000	171	0.109	0	0.000	0	0.000
Nullabgleich	0	0.000	0	0.000	71	0.045	0	0.000	0	0.000
Dislokation	0	0.000	0	0.000	0	0.000	0	0.000	57	0.101
Leitung abgeknickt	0	0.000	0	0.000	54	0.034	0	0.000	0	0.000
EKG zählt falsch	0	0.000	43	0.741	0	0.000	0	0.000	0	0.000
Elektrode ab	1	0.005	6	0.103	0	0.000	0	0.000	0	0.000
Fehlerhafte Platzierung	0	0.000	0	0.000	0	0.000	3	0.375	0	0.000
Niedervoltage	1	0.005	0	0.000	0	0.000	0	0.000	0	0.000
Perfusionsstörung	0	0.000	0	0.000	0	0.000	0	0.000	1	0.002

Tabelle B.10: Ursachen für technisch falsche Alarme für die einzelnen Alarmquellen. Bei einem Alarm können mehrere Ursachen gleichzeitig auftreten.

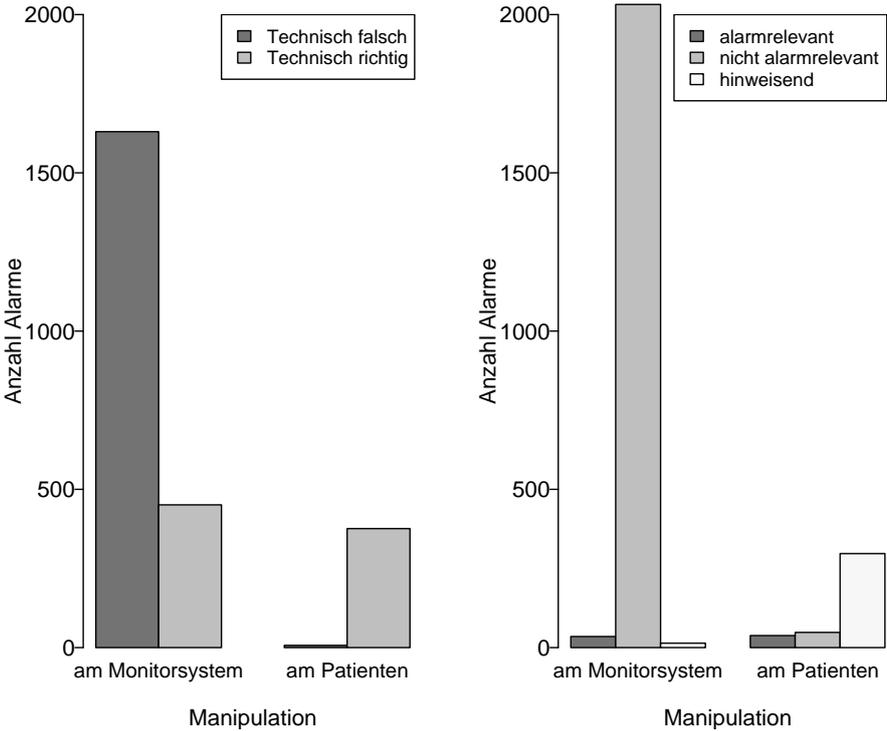


Abbildung B.1: Manipulationen am Patienten und am Monitorsystem

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe
		ja	nein	hinw.	
richtig	nein	344	149	1688	2181
		0.16	0.07	0.77	1.00
		0.90	0.77	0.85	0.85
		0.08	0.04	0.41	0.53
	ja	37	44	290	371
		0.10	0.12	0.78	1.00
		0.10	0.23	0.15	0.15
		0.01	0.01	0.07	0.09
	Summe	381	193	1978	2552
		0.15	0.08	0.78	1.00
		0.52	0.14	1.00	0.63
		0.09	0.05	0.48	0.62
falsch	nein	331	261	-	592
		0.56	0.44	-	1.00
		0.94	0.23	-	0.39
		0.08	0.06	-	0.14
	ja	20	893	-	913
		0.02	0.98	-	1.00
		0.06	0.77	-	0.61
		0.00	0.22	-	0.22
	Summe	351	1154	-	1505
		0.23	0.77	-	1.00
		0.48	0.86	-	0.37
		0.08	0.28	-	0.36
Insgesamt	nein	675	410	1688	2773
		0.24	0.15	0.61	1.00
		0.92	0.30	0.85	0.68
		0.16	0.10	0.41	0.67
	ja	57	937	290	1284
		0.04	0.73	0.23	1.00
		0.08	0.70	0.15	0.32
		0.01	0.23	0.07	0.31
	Summe	732	1347	1978	4057
		0.18	0.33	0.49	1.00
		1.00	1.00	1.00	1.00
		0.18	0.32	0.48	0.98
Nicht zu beurteilen				47	
Nicht bewertet				42	
Alarme insgesamt				4146	

Tabelle B.11: Alarm-Annotationen der ausgelösten Schwellwertalarml. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe	
		ja	nein	hinw.		
richtig	nein	0	3	11	14	
		0.00	0.21	0.79	1.00	
		-	0.75	0.65	0.67	
		0.00	0.00	0.01	0.02	
		<hr/>				
		0	1	6	7	
	0.00	0.14	0.86	1.00		
	-	0.25	0.35	0.33		
	0.00	0.00	0.01	0.01		
	<hr/>					
	Summe	0	4	17	21	
		0.00	0.19	0.81	1.00	
	0.00	0.01	1.00	0.02		
	0.00	0.00	0.02	0.02		
falsch	nein	73	80	-	153	
		0.48	0.52	-	1.00	
		0.88	0.11	-	0.18	
		0.08	0.09	-	0.18	
		<hr/>				
		10	679	-	689	
	0.01	0.99	-	1.00		
	0.12	0.89	-	0.82		
	0.01	0.78	-	0.79		
	<hr/>					
	Summe	83	759	-	842	
		0.10	0.90	-	1.00	
	1.00	0.99	-	0.98		
	0.10	0.87	-	0.96		
Insgesamt	nein	73	83	11	167	
		0.44	0.50	0.07	1.00	
		0.88	0.11	0.65	0.19	
		0.08	0.10	0.01	0.19	
		<hr/>				
		10	680	6	696	
	0.01	0.98	0.01	1.00		
	0.12	0.89	0.35	0.81		
	0.01	0.78	0.01	0.80		
	<hr/>					
	Summe	83	763	17	863	
		0.10	0.88	0.02	1.00	
	1.00	1.00	1.00	1.00		
	0.10	0.87	0.02	0.99		
<hr/>						
Nicht zu beurteilen				2		
Nicht bewertet				8		
Alarme insgesamt				873		

Tabelle B.12: Alarm-Annotationen der ausgelösten sonstigen ernststen Alarme. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe
		ja	nein	hinw.	
richtig	nein	42	132	156	330
		0.13	0.40	0.47	1.00
		0.91	0.24	0.91	0.44
		0.05	0.16	0.19	0.41
	ja	4	408	15	427
		0.01	0.96	0.04	1.00
		0.09	0.76	0.09	0.56
		0.00	0.51	0.02	0.53
	Summe	46	540	171	757
		0.06	0.71	0.23	1.00
		0.81	0.95	1.00	0.95
		0.06	0.67	0.21	0.94
falsch	nein	10	5	-	15
		0.67	0.33	-	1.00
		0.91	0.18	-	0.38
		0.01	0.01	-	0.02
	ja	1	23	-	24
		0.04	0.96	-	1.00
		0.09	0.82	-	0.62
		0.00	0.03	-	0.03
	Summe	11	28	-	39
		0.28	0.72	-	1.00
		0.19	0.05	-	0.05
		0.01	0.03	-	0.05
Insgesamt	nein	52	137	156	345
		0.15	0.40	0.45	1.00
		0.91	0.24	0.91	0.43
		0.06	0.17	0.19	0.43
	ja	5	431	15	451
		0.01	0.96	0.03	1.00
		0.09	0.76	0.09	0.57
		0.01	0.53	0.02	0.56
	Summe	57	568	171	796
		0.07	0.71	0.21	1.00
		1.00	1.00	1.00	1.00
		0.07	0.70	0.21	0.99
Nicht zu beurteilen				6	
Nicht bewertet				5	
Alarmer insgesamt				807	

Tabelle B.13: Alarm-Annotationen der ausgelösten technischen Alarme. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe
		ja	nein	hinw.	
richtig	nein	12	1	57	70
		0.17	0.01	0.81	1.00
		0.92	0.05	1.00	0.76
		0.11	0.01	0.53	0.65
	ja	1	21	0	22
		0.05	0.95	0.00	1.00
		0.08	0.95	0.00	0.24
		0.01	0.19	0.00	0.20
	Summe	13	22	57	92
		0.14	0.24	0.62	1.00
		1.00	0.65	1.00	0.88
		0.12	0.20	0.53	0.85
falsch	nein	0	0	-	0
		-	-	-	-
		-	0.00	-	0.00
		0.00	0.00	-	0.00
	ja	0	12	-	12
		0.00	1.00	-	1.00
		-	1.00	-	1.00
		0.00	0.11	-	0.11
	Summe	0	12	-	12
		0.00	1.00	-	1.00
		0.00	0.35	-	0.12
		0.00	0.11	-	0.11
Insgesamt	nein	12	1	57	70
		0.17	0.01	0.81	1.00
		0.92	0.03	1.00	0.67
		0.11	0.01	0.53	0.65
	ja	1	33	0	34
		0.03	0.97	0.00	1.00
		0.08	0.97	0.00	0.33
		0.01	0.31	0.00	0.31
	Summe	13	34	57	104
		0.12	0.33	0.55	1.00
		1.00	1.00	1.00	1.00
		0.12	0.31	0.53	0.96
Nicht zu beurteilen				0	
Nicht bewertet				4	
Alarmer insgesamt				108	

Tabelle B.14: Alarm-Annotationen der ausgelösten Arrhythmie-Alarmer. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarmer an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarmer an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe
		ja	nein	hinw.	
richtig	nein	213	85	987	1285
		0.17	0.07	0.77	1.00
		0.88	0.71	0.81	0.81
		0.09	0.04	0.42	0.55
	ja	29	35	235	299
		0.10	0.12	0.79	1.00
		0.12	0.29	0.19	0.19
		0.01	0.01	0.10	0.13
	Summe	242	120	1222	1584
		0.15	0.08	0.77	1.00
		0.60	0.18	1.00	0.69
		0.10	0.05	0.52	0.68
	falsch	nein	150	50	-
0.75			0.25	-	1.00
0.91			0.09	-	0.28
0.06			0.02	-	0.09
ja		14	501	-	515
		0.03	0.97	-	1.00
		0.09	0.91	-	0.72
		0.01	0.21	-	0.22
Summe		164	551	-	715
		0.23	0.77	-	1.00
		0.40	0.82	-	0.31
		0.07	0.24	-	0.31
Insgesamt		nein	363	135	987
	0.24		0.09	0.66	1.00
	0.89		0.20	0.81	0.65
	0.15		0.06	0.42	0.63
	ja	43	536	235	814
		0.05	0.66	0.29	1.00
		0.11	0.80	0.19	0.35
		0.02	0.23	0.10	0.35
	Summe	406	671	1222	2299
		0.18	0.29	0.53	1.00
		1.00	1.00	1.00	1.00
		0.17	0.29	0.52	0.98
	Nicht zu beurteilen				17
Nicht bewertet				26	
Alarme insgesamt				2342	

Tabelle B.15: Alarm-Annotationen der ausgelösten Alarme für systolischen und mittleren arteriellen Blutdruck. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe
		ja	nein	hinw.	
richtig	nein	95	5	100	200
		0.47	0.03	0.50	1.00
		0.96	0.56	0.93	0.93
		0.12	0.01	0.13	0.26
	ja	4	4	8	16
		0.25	0.25	0.50	1.00
		0.04	0.44	0.07	0.07
		0.01	0.01	0.01	0.02
	Summe	99	9	108	216
		0.46	0.04	0.50	1.00
		0.37	0.02	1.00	0.29
		0.13	0.01	0.14	0.28
falsch	nein	168	72	-	240
		0.70	0.30	-	1.00
		0.98	0.20	-	0.45
		0.21	0.09	-	0.31
	ja	4	287	-	291
		0.01	0.99	-	1.00
		0.02	0.80	-	0.55
		0.01	0.37	-	0.37
	Summe	172	359	-	531
		0.32	0.68	-	1.00
		0.63	0.98	-	0.71
		0.22	0.46	-	0.68
Insgesamt	nein	263	77	100	440
		0.60	0.17	0.23	1.00
		0.97	0.21	0.93	0.59
		0.34	0.10	0.13	0.56
	ja	8	291	8	307
		0.03	0.95	0.03	1.00
		0.03	0.79	0.07	0.41
		0.01	0.37	0.01	0.39
	Summe	271	368	108	747
		0.36	0.49	0.14	1.00
		1.00	1.00	1.00	1.00
		0.35	0.47	0.14	0.95
Nicht zu beurteilen				29	
Nicht bewertet				8	
Alarmer insgesamt				784	

Tabelle B.16: Alarm-Annotationen der ausgelösten Alarme für die Sauerstoffsättigung. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe
		ja	nein	hinw.	
richtig	nein	27	59	550	636
		0.04	0.09	0.86	1.00
		0.90	0.92	0.95	0.95
		0.04	0.08	0.76	0.87
	ja	3	5	26	34
		0.09	0.15	0.76	1.00
		0.10	0.08	0.05	0.05
		0.00	0.01	0.04	0.05
	Summe	30	64	576	670
		0.04	0.10	0.86	1.00
		1.00	0.55	1.00	0.93
		0.04	0.09	0.79	0.92
falsch	nein	0	22	-	22
		0.00	1.00	-	1.00
		-	0.42	-	0.42
		0.00	0.03	-	0.03
	ja	0	30	-	30
		0.00	1.00	-	1.00
		-	0.58	-	0.58
		0.00	0.04	-	0.04
	Summe	0	52	-	52
		0.00	1.00	-	1.00
		0.00	0.45	-	0.07
		0.00	0.07	-	0.07
Insgesamt	nein	27	81	550	658
		0.04	0.12	0.84	1.00
		0.90	0.70	0.95	0.91
		0.04	0.11	0.76	0.91
	ja	3	35	26	64
		0.05	0.55	0.41	1.00
		0.10	0.30	0.05	0.09
		0.00	0.05	0.04	0.09
	Summe	30	116	576	722
		0.04	0.16	0.80	1.00
		1.00	1.00	1.00	1.00
		0.04	0.16	0.79	0.99
Nicht zu beurteilen				0	
Nicht bewertet				5	
Alarmer insgesamt				727	

Tabelle B.17: Alarm-Annotationen der ausgelösten Alarme für die Herzfrequenz. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Technische Validität	Manipulation	Alarmrelevanter Zustand			Summe	
		ja	nein	hinw.		
richtig	nein	0	1	5	6	
		0.00	0.17	0.83	1.00	
		-	1.00	1.00	1.00	
		0.00	0.01	0.03	0.03	
		0	0	0	0	
		-	-	-	-	
	ja	-	0.00	0.00	0.00	
		0.00	0.00	0.00	0.00	
		Summe	0	1	5	6
		0.00	0.17	0.83	1.00	
		0.00	0.01	1.00	0.03	
		0.00	0.01	0.03	0.03	
falsch	nein	2	146	-	148	
		0.01	0.99	-	1.00	
		0.67	0.80	-	0.80	
		0.01	0.74	-	0.76	
		1	36	-	37	
		0.03	0.97	-	1.00	
	ja	0.33	0.20	-	0.20	
		0.01	0.18	-	0.19	
		Summe	3	182	-	185
		0.02	0.98	-	1.00	
		1.00	0.99	-	0.97	
		0.02	0.93	-	0.94	
Insgesamt	nein	2	147	5	154	
		0.01	0.95	0.03	1.00	
		0.67	0.80	1.00	0.81	
		0.01	0.75	0.03	0.79	
		1	36	0	37	
		0.03	0.97	0.00	1.00	
	ja	0.33	0.20	0.00	0.19	
		0.01	0.18	0.00	0.19	
		Summe	3	183	5	191
		0.02	0.96	0.03	1.00	
		1.00	1.00	1.00	1.00	
		0.02	0.93	0.03	0.97	
Nicht zu beurteilen				0		
Nicht bewertet				5		
Alarmer insgesamt				196		

Tabelle B.18: Alarm-Annotationen der ausgelösten Alarme für die Atemfrequenz. Die erste Zeile eines Blockes gibt dabei immer die absolute Häufigkeit, die zweite Zeile den Anteil der Alarme an der entsprechenden Zeilensumme, die dritte Zeile den Anteil der Alarme an der entsprechenden Spalten(-zwischen-)summe und die vierte Zeile den Anteil an allen Alarmen an.

Anhang C

Ergänzungen zum Alarmsystem-Vergleich

Fall	Dauer	ART.S/ART.M			HR			SpO2		
		A	A/Std	AR	A	A/Std	AR	A	A/Std	AR
3	14:28:14	0	0.00	0	0	0.00	0	7	0.48	1
6	09:13:07	58	6.29	16	1	0.11	0	1	0.11	0
7	24:10:43	9	0.37	2	1	0.04	1	1	0.04	0
11	08:04:22	121	14.99	6	20	2.48	0	43	5.33	7
12	09:21:30	4	0.43	0	0	0.00	0	9	0.96	1
16	20:27:28	141	6.89	59	3	0.15	0	114	5.57	79
19	06:53:12	32	4.65	3	1	0.15	0	4	0.58	0
20	04:09:33	1	0.24	0	23	5.53	0	7	1.68	7
25	24:00:00	22	0.92	10	0	0.00	0	5	0.21	0
27	24:42:01	44	1.78	1	2	0.08	0	32	1.30	4
30	11:44:41	62	5.28	5	0	0.00	0	13	1.11	0
31	22:42:54	71	3.13	3	30	1.32	0	6	0.26	6
32	22:51:51	99	4.33	8	1	0.04	0	35	1.53	4
33	06:18:20	5	0.79	2	0	0.00	0	1	0.16	0
34	08:07:21	56	6.89	8	6	0.74	3	25	3.08	6
35	05:09:55	3	0.58	0	0	0.00	0	1	0.19	0
38	16:21:59	5	0.31	1	3	0.18	0	9	0.55	5
41	17:24:16	5	0.29	0	0	0.00	0	0	0.00	0
58	15:35:45	23	1.47	4	0	0.00	0	8	0.51	4
59	06:00:48	8	1.33	2	3	0.50	3	0	0.00	0
64	23:24:15	18	0.77	1	376	16.07	2	14	0.60	4
65	14:16:41	73	5.11	1	11	0.77	0	22	1.54	3
67	19:06:04	5	0.26	1	0	0.00	0	8	0.42	5
Insgesamt	334:35:00	865	2.59	133	481	1.44	9	365	1.09	136

Tabelle C.1: Ausgewählte Datensätze mit Beobachtungsdauer, Anzahl Schwellwertalarme (A), Anzahl Schwellwertalarme pro Stunde (A/Std) sowie Anzahl alarmrelevant annotierter Alarme (AR) für arterielle Blutdrücke, Herzfrequenz und Sauerstoffsättigung

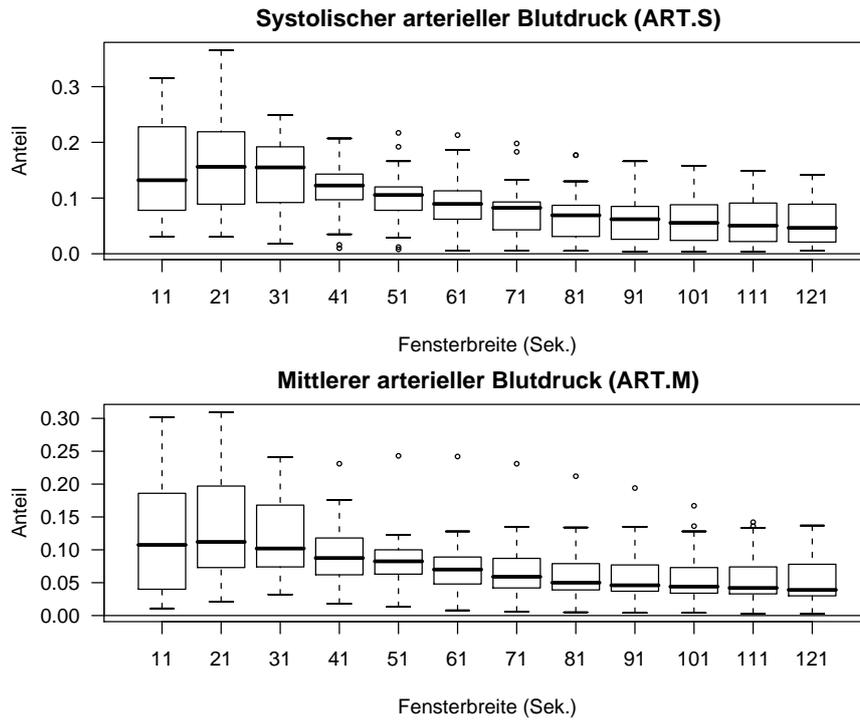


Abbildung C.1: Anteil der Fenster für die Fälle der Teilstichprobe, in denen der geschätzte Signalwert größer als das Maximum bzw. kleiner als das Minimum der Originalwerte im Fenster ist, getrennt für den systolischen und mittleren arteriellen Blutdruck

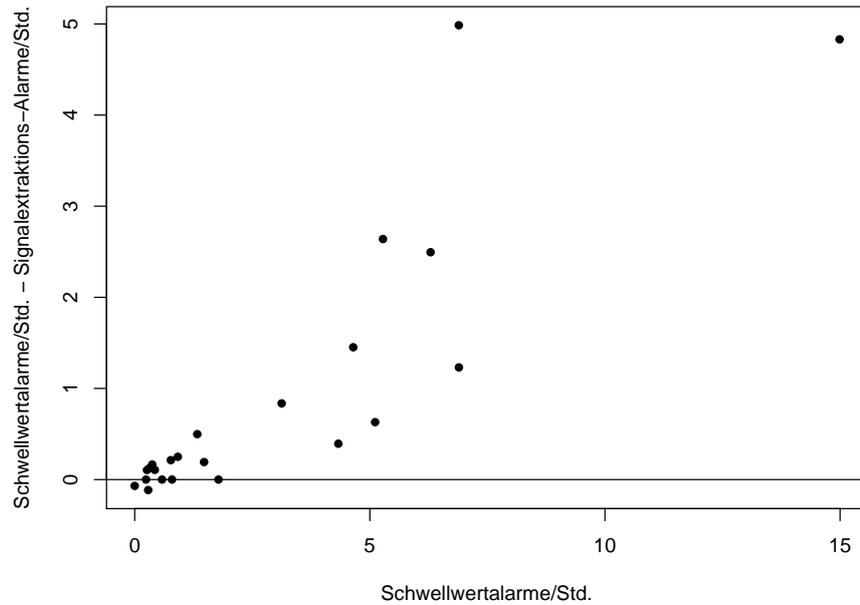


Abbildung C.2: Anzahl Schwellwertalarme pro Stunde gegen die Differenz von Anzahl Schwellwertalarme pro Stunde und Anzahl Signalextraktions-Alarme pro Stunde für die arteriellen Blutdrücke (hier beispielhaft für Fensterbreite 91 Sek.)

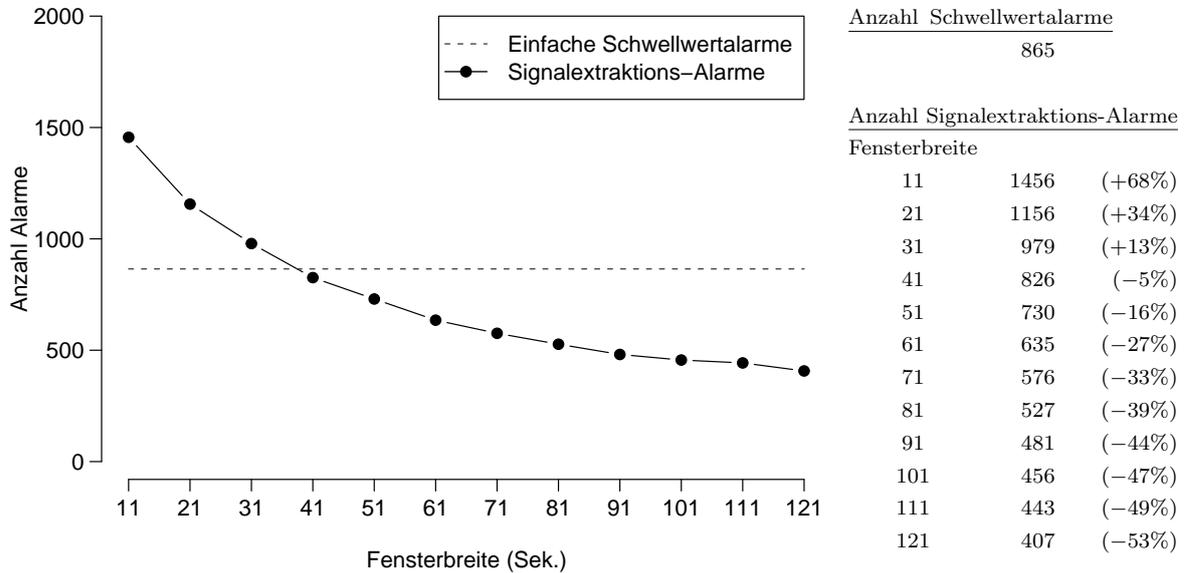


Abbildung C.4: Zusatzregel: Alarm dauert an, bis extrahiertes Signal wieder mehr als eine Einheit (1 mmHg) innerhalb der Grenzen liegt; Anzahl neuer Alarme für den systolischen und mittleren arteriellen Blutdruck im Vergleich zur Anzahl einfacher Schwellwertalarmer

Neues System erzeugt	Fensterbreite (Sek.)											
	11	21	31	41	51	61	71	81	91	101	111	121
mehr Alarme	22	19	18	14	10	6	4	3	1	1	1	1
gleich viele Alarme	0	2	0	2	4	2	3	3	4	5	4	3
weniger Alarme	1	2	5	7	9	15	16	17	18	17	18	19

Tabelle C.3: Zusatzregel: Alarm dauert an, bis extrahiertes Signal wieder mehr als eine Einheit (1 mmHg) innerhalb der Grenzen liegt; Anzahl Fälle, in denen das neue Signalextraktions-Alarmsystem mehr, gleich viele und weniger Alarme als das einfache Schwellwertalarm-system produziert

		Fensterbreite											
	Sek.	11	21	31	41	51	61	71	81	91	101	111	121
Ingesamt	±5	100	97.0	85.0	69.9	60.2	60.2	56.4	57.1	54.9	54.1	56.4	54.9
	±10	100	99.2	95.5	85.7	76.7	69.2	63.9	61.7	59.4	57.1	60.9	58.6
	±30	100	100	97.0	91.7	84.2	78.9	75.9	72.9	70.7	69.2	69.9	68.4
	±60	100	100	97.7	94.7	90.2	83.5	80.5	78.9	78.2	75.2	76.7	75.2
	±120	100	100	99.2	97.0	92.5	85.7	85.0	82.7	82.7	81.2	82.0	79.7
	±180	100	100	99.2	97.7	94.7	88.7	89.5	86.5	86.5	86.5	87.2	85.7
	Technisch falsch	±5	100	100	83.6	58.9	49.3	49.3	46.6	47.9	43.8	42.5	43.8
±10		100	100	97.3	82.2	71.2	61.6	52.1	52.1	50.7	45.2	47.9	46.6
±30		100	100	97.3	87.7	76.7	67.1	64.4	63.0	58.9	56.2	57.5	57.5
±60		100	100	98.6	93.2	84.9	72.6	71.2	69.9	67.1	64.4	64.4	64.4
±120		100	100	100	97.3	89.0	76.7	76.7	74.0	74.0	71.2	71.2	69.9
±180		100	100	100	98.6	93.2	82.2	83.6	79.5	79.5	79.5	79.5	79.5
Technisch richtig		±5	100	93.3	86.7	83.3	73.3	73.3	68.3	68.3	68.3	68.3	71.7
	±10	100	98.3	93.3	90.0	83.3	78.3	78.3	73.3	70.0	71.7	76.7	73.3
	±30	100	100	96.7	96.7	93.3	93.3	90.0	85.0	85.0	85.0	85.0	81.7
	±60	100	100	96.7	96.7	96.7	96.7	91.7	90.0	91.7	88.3	91.7	88.3
	±120	100	100	98.3	96.7	96.7	96.7	95.0	93.3	93.3	93.3	95.0	91.7
	±180	100	100	98.3	96.7	96.7	96.7	96.7	95.0	95.0	95.0	96.7	93.3

Tabelle C.4: Zusatzregel: Alarm dauert an, bis extrahiertes Signal wieder mehr als eine Einheit (1 mmHg) innerhalb der Grenzen liegt; Sensitivitäten (in %) für verschiedene Fensterbreiten und Intervalle um einen als alarmrelevant annotierten Schwellwertalarm

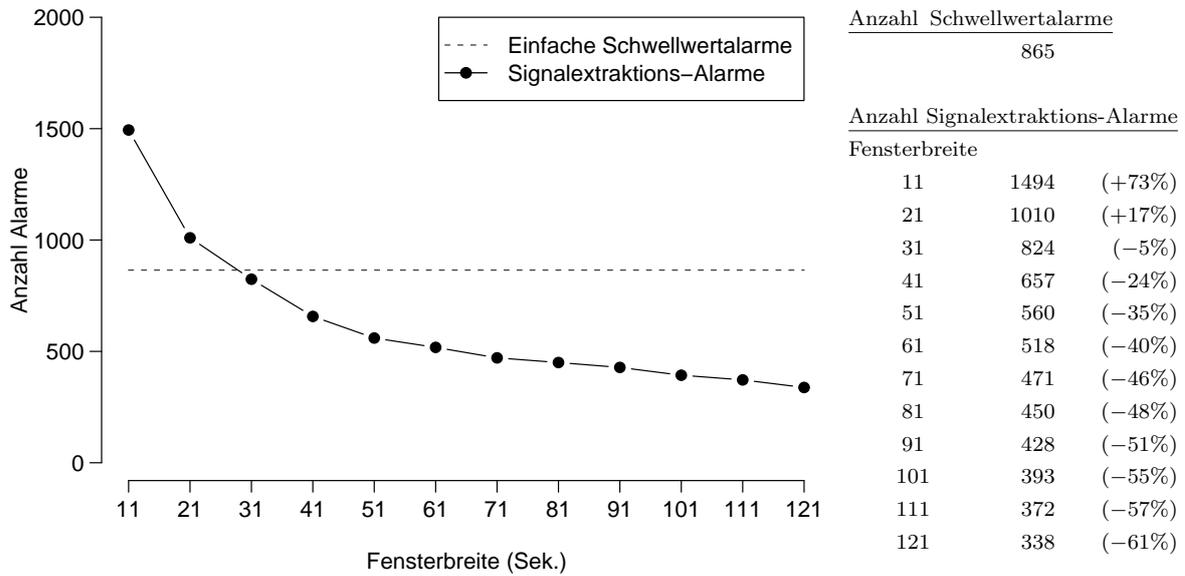


Abbildung C.5: Schätzung für die Fenstermitte: Anzahl neuer Alarme für den systolischen und mittleren arteriellen Blutdruck im Vergleich zur Anzahl einfacher Schwellwertalarme

	Sek.	Fensterbreite											
		11	21	31	41	51	61	71	81	91	101	111	121
Insge- samt	±5	99.2	57.9	38.3	31.6	31.6	27.8	26.3	24.8	24.1	24.1	24.1	24.1
	±10	99.2	79.7	58.6	44.4	34.6	32.3	27.8	25.6	25.6	24.1	25.6	24.8
	±30	99.2	86.5	78.2	72.2	66.9	62.4	51.9	40.6	36.8	33.1	30.1	27.8
	±60	99.2	90.2	83.5	78.9	73.7	72.9	65.4	60.9	62.4	54.1	52.6	47.4
	±120	99.2	91.0	86.5	82.0	76.7	77.4	72.2	69.2	69.2	64.7	63.9	59.4
	±180	99.2	94.0	87.2	83.5	78.2	78.9	74.4	72.9	72.2	69.2	68.4	63.2
	Tech- nisch falsch	±5	100	47.9	34.2	30.1	30.1	30.1	30.1	28.8	27.4	27.4	28.8
±10		100	71.2	52.1	39.7	31.5	32.9	30.1	30.1	30.1	27.4	30.1	30.1
±30		100	79.5	71.2	58.9	54.8	52.1	43.8	37.0	35.6	32.9	31.5	31.5
±60		100	84.9	76.7	67.1	58.9	61.6	53.4	52.1	54.8	45.2	46.6	43.8
±120		100	86.3	79.5	71.2	63.0	64.4	57.5	56.2	60.3	56.2	56.2	54.8
±180		100	91.8	80.8	74.0	64.4	67.1	58.9	57.5	61.6	57.5	57.5	56.2
Tech- nisch richtig		±5	98.3	70.0	43.3	33.3	33.3	25.0	21.7	20.0	20.0	20.0	18.3
	±10	98.3	90.0	66.7	50.0	38.3	31.7	25.0	20.0	20.0	20.0	20.0	18.3
	±30	98.3	95.0	86.7	88.3	81.7	75.0	61.7	45.0	38.3	33.3	28.3	23.3
	±60	98.3	96.7	91.7	93.3	91.7	86.7	80.0	71.7	71.7	65.0	60.0	51.7
	±120	98.3	96.7	95.0	95.0	93.3	93.3	90.0	85.0	80.0	75.0	73.3	65.0
	±180	98.3	96.7	95.0	95.0	95.0	93.3	93.3	91.7	85.0	83.3	81.7	71.7

Tabelle C.5: Schätzung für die Fenstermitte: Sensitivitäten (in %) für verschiedene Fensterbreiten und Intervalle um einen als alarmrelevant annotierten Schwellwertalarm

Bezeichnung	Anzahl	Beschreibung						
Gleichzeitige Alarmer ($A_{alt}-A_{neu}$)	712	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	6	15	37.52	35.25	779
Unterdrückte Alarmer ($A_{alt}-\overline{A_{neu}}$)	153	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	3	4	5.82	7	21
Alarmrelevant: technisch richtig: 5, technisch falsch: 20								
Gleichzeitige Alarmer ($A_{neu}-A_{alt}$)	374	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	36	67	119.3	139	1296
		Delay (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			-190	-1	5	3.75	15.75	250
Zusätzliche Alarmer ($A_{neu}-\overline{A_{alt}}$)	107	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	13	28	42.93	47	228

Tabelle C.6: Auswertung nach der Methode aus Charbonnier, Gentil (2007). Vergleich zwischen alten und neuen Alarmen (Fensterbreite 91) für die arteriellen Blutdrücke.

Bezeichnung	Anzahl	Beschreibung						
Gleichzeitige Alarmer ($A_{alt}-A_{neu}$)	438	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			2	3	5	8.45	9	226
Unterdrückte Alarmer ($A_{alt}-\overline{A_{neu}}$)	43	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			2	3	3	3.67	5	8
Alarmrelevant: technisch richtig: 0, technisch falsch: 0								
Gleichzeitige Alarmer ($A_{neu}-A_{alt}$)	343	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	9	13	18.97	18	408
		Delay (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			-47	1.5	6	8.61	15	45
Zusätzliche Alarmer ($A_{neu}-\overline{A_{alt}}$)	63	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	6	10	12.6	13	79

Tabelle C.7: Auswertung nach der Methode aus Charbonnier, Gentil (2007). Vergleich zwischen alten und neuen Alarmen (Fensterbreite 31) für die Herzfrequenz.

Bezeichnung	Anzahl	Beschreibung						
Gleichzeitige Alarmer ($A_{alt}-A_{neu}$)	251	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	7	19	41.91	42	459
Unterdrückte Alarmer ($A_{alt}-\overline{A_{neu}}$)	114	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	3	5.5	6.61	9	20
Alarmrelevant: technisch richtig: 4, technisch falsch: 30								
Gleichzeitige Alarmer ($A_{neu}-A_{alt}$)	156	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			5	47.25	100	196	246	1978
		Delay (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			-338	-5.25	6.5	-3.19	20.25	268
Zusätzliche Alarmer ($A_{neu}-\overline{A_{alt}}$)	62	Dauer (Sek.)	Min.	1.Q.	Median	Mittel	3.Q.	Max.
			1	22	54	74.84	114	303

Tabelle C.8: Auswertung nach der Methode aus Charbonnier, Gentil (2007). Vergleich zwischen alten und neuen Alarmen (Fensterbreite 91) für die Sauerstoffsättigung.

Fall	Dauer	ART.S/ART.M			HR			SpO2		
		A	A/Std	AR	A	A/Std	AR	A	A/Std	AR
1	03:00:01	19	6.33	4	0	0.00	0	5	1.67	2
2	10:44:53	41	3.81	22	0	0.00	0	2	0.19	2
4	20:14:44	6	0.30	0	9	0.44	1	2	0.10	0
5	00:55:01	2	2.18	0	0	0.00	0	0	0.00	0
8	19:56:10	27	1.35	10	1	0.05	0	1	0.05	1
9	14:33:37	190	13.05	36	31	2.13	7	50	3.43	25
10	14:20:57	138	9.62	17	18	1.25	5	53	3.69	34
13	02:52:05	1	0.35	0	0	0.00	0	1	0.35	0
14	13:47:54	25	1.81	6	0	0.00	0	38	2.75	19
15	08:54:52	32	3.59	8	0	0.00	0	0	0.00	0
17	05:13:00	9	1.73	5	0	0.00	0	5	0.96	3
18	17:51:56	27	1.51	4	0	0.00	0	8	0.45	2
21	05:29:30	12	2.19	2	21	3.82	2	6	1.09	0
22	05:55:28	13	2.19	6	0	0.00	0	2	0.34	0
23	11:13:46	12	1.07	1	0	0.00	0	15	1.34	0
24	22:18:10	20	0.90	4	1	0.04	0	1	0.04	1
26	16:31:44	9	0.54	0	9	0.54	0	2	0.12	0
28	23:56:34	82	3.42	0	4	0.17	1	14	0.58	4
29	24:56:54	32	1.28	1	50	2.00	0	25	1.00	0
36	15:41:55	30	1.91	9	6	0.38	0	18	1.15	3
37	14:58:23	21	1.40	5	2	0.13	0	18	1.20	3
39	17:18:07	5	0.29	1	0	0.00	0	5	0.29	3
40	22:13:45	38	1.71	7	1	0.04	1	4	0.18	2
42	19:23:53	13	0.67	0	0	0.00	0	1	0.05	0
43	24:11:03	7	0.29	0	0	0.00	0	8	0.33	0
44	22:57:58	70	3.05	7	2	0.09	0	4	0.17	0
45	10:49:26	11	1.02	3	0	0.00	0	9	0.83	2
46	25:15:48	80	3.17	17	0	0.00	0	9	0.36	0
47	25:17:15	16	0.63	2	3	0.12	0	0	0.00	0
48	17:34:58	19	1.08	2	15	0.85	2	10	0.57	2
49	01:16:40	1	0.78	0	0	0.00	0	0	0.00	0
50	09:09:50	11	1.20	4	1	0.11	0	6	0.65	3
51	11:40:28	20	1.71	3	6	0.51	1	13	1.11	2
52	24:00:00	3	0.12	0	0	0.00	0	2	0.08	0
53	07:15:53	24	3.30	2	5	0.69	1	1	0.14	0
54	01:19:01	0	0.00	0	0	0.00	0	0	0.00	0
55	18:38:05	85	4.56	11	1	0.05	0	31	1.66	10
56	07:10:41	9	1.25	0	10	1.39	0	1	0.14	0
57	23:38:21	94	3.98	48	0	0.00	0	8	0.34	3
60	08:08:50	22	2.70	6	3	0.37	0	0	0.00	0
61	01:58:42	5	2.53	1	0	0.00	0	2	1.01	1
62	19:45:00	38	1.92	0	44	2.23	0	26	1.32	2
63	21:43:31	98	4.51	12	2	0.09	0	3	0.14	1
66	09:07:59	4	0.44	1	0	0.00	0	4	0.44	4
68	24:14:32	56	2.31	6	1	0.04	0	6	0.25	1
Insgesamt	647:37:20	1477	2.28	273	246	0.38	21	419	0.65	135

Tabelle C.9: Für die Validierung ausgewählte Datensätze mit Beobachtungsdauer, Anzahl Schwellwertalarme (A), Anzahl Schwellwertalarme pro Stunde (A/Std) sowie Anzahl alarmrelevant annotierter Alarme (AR) für ART.M/ART.S, HR und SpO2

Fall	Anzahl AR (TR)	Anzahl erkannt	$\hat{S}e(A_{neu})$	falsch pos. alt	falsch pos. neu	FP_{max}	$\hat{S}p(A_{alt})$	$\hat{S}p(A_{neu})$
1	3	3	100	13	10	37	64.9	73.0
2	21	21	100	18	10	137	86.9	92.7
4	0	0	-	6	1	300	98.0	99.7
5	0	0	-	2	3	11	81.8	72.7
8	5	4	80.0	15	13	279	94.6	95.3
9	10	8	80.0	138	50	180	23.3	72.2
10	17	13	76.5	107	45	188	43.1	76.1
13	0	0	-	1	1	43	97.7	97.7
14	0	0	-	19	10	188	89.9	94.7
15	6	6	100	24	18	121	80.2	85.1
17	0	0	-	4	2	51	92.2	96.1
18	4	4	100	21	12	249	91.6	95.2
21	2	1	50.0	9	4	77	88.3	94.8
22	6	5	83.3	7	5	81	91.4	93.8
23	1	1	100	11	5	155	92.9	96.8
24	4	4	100	14	5	322	95.7	98.4
26	0	0	-	9	6	229	96.1	97.4
28	0	0	-	82	43	328	75.0	86.9
29	1	1	100	30	26	326	90.8	92.0
36	6	6	100	18	19	199	91.0	90.5
37	5	5	100	12	7	202	94.1	96.5
39	1	1	100	3	1	252	98.8	99.6
40	3	3	100	29	13	313	90.7	95.8
42	0	0	-	13	5	283	95.4	98.2
43	0	0	-	7	4	352	98.0	98.9
44	7	6	85.7	59	31	297	80.1	89.6
45	3	3	100	7	5	145	95.2	96.6
46	17	16	94.1	62	56	326	81.0	82.8
47	2	1	50.0	12	11	357	96.6	96.9
48	2	2	100	17	15	242	93.0	93.8
49	0	0	-	1	1	18	94.4	94.4
50	4	4	100	7	6	123	94.3	95.1
51	0	0	-	16	3	164	90.2	98.2
52	0	0	-	3	1	341	99.1	99.7
53	2	2	100	20	10	103	80.6	90.3
54	0	0	-	0	0	9	100	100
55	1	0	0	72	31	259	72.2	88.0
56	0	0	-	9	10	99	90.9	89.9
57	35	31	88.6	41	25	307	86.6	91.9
60	6	6	100	13	7	111	88.3	93.7
61	0	0	-	4	2	28	85.7	92.9
62	0	0	-	38	20	270	85.9	92.6
63	5	5	100	84	32	291	71.1	89.0
66	1	1	100	3	4	129	97.7	96.9
68	2	2	100	50	38	337	85.2	88.7
Insg.	182	165	90.7	1130	626	8859	87.2	92.9

Tabelle C.10: Anzahl alarmrelevanter Zustände bei technisch richtiger Datenaufzeichnung für die arteriellen Blutdrücke, Anzahl der vom neuen System erkannten alarmrelevanten Zustände, geschätzte Sensitivität (in %) für den neuen Alarm-Algorithmus, Anzahl falsch positiver Alarmer, „Worst-case“-Anzahl und geschätzte Spezifitäten (in %) getrennt für die einzelnen Fälle und insgesamt.

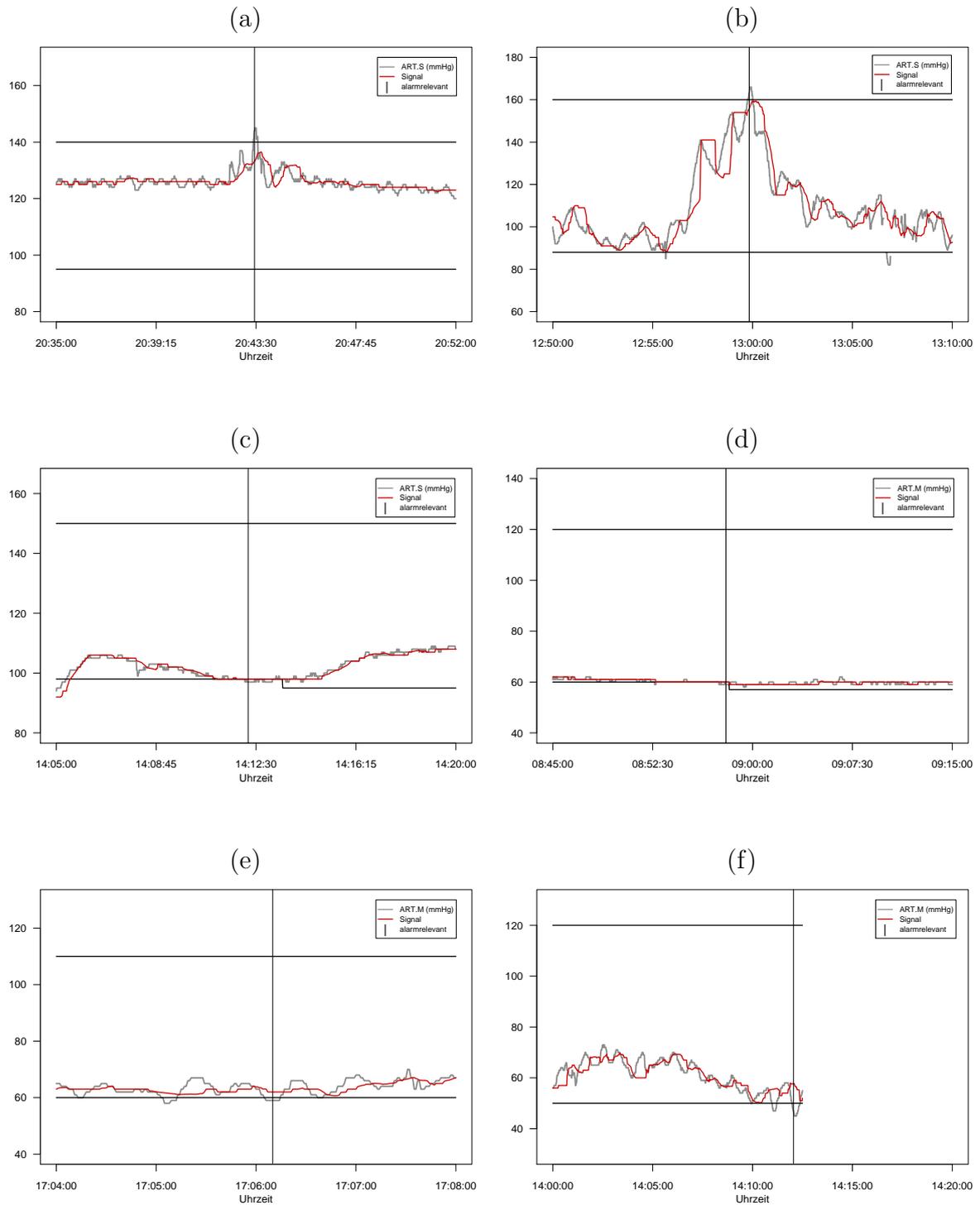


Abbildung C.6: Nicht erkannte alarmrelevante Situationen für die arteriellen Blutdrücke

Fall	Anzahl AR (TR)	Anzahl erkannt	$\hat{S}e(A_{neu})$	falsch pos. alt	falsch pos. neu	FP_{max}	$\hat{S}p(A_{alt})$	$\hat{S}p(A_{neu})$
1	0	0	-	0	0	15	100	100
2	0	0	-	0	0	57	100	100
4	1	0	0	8	4	120	93.3	96.7
5	0	0	-	0	0	4	100	100
8	0	0	-	1	1	113	99.1	99.1
9	7	4	57.1	24	38	74	67.6	48.6
10	5	1	20.0	12	10	78	84.6	87.2
13	0	0	-	0	0	17	100	100
14	0	0	-	0	0	76	100	100
15	0	0	-	0	0	50	100	100
17	0	0	-	0	0	30	100	100
18	0	0	-	0	0	100	100	100
21	2	1	50.0	19	13	31	38.7	58.1
22	0	0	-	0	0	33	100	100
23	0	0	-	0	0	62	100	100
24	0	0	-	1	1	129	99.2	99.2
26	0	0	-	9	6	92	90.2	93.5
28	1	0	0	3	3	131	97.7	97.7
29	0	0	-	50	5	131	61.8	96.2
36	0	0	-	6	11	81	92.6	86.4
37	0	0	-	2	1	81	97.5	98.8
39	0	0	-	0	1	101	100	99.0
40	1	1	100	0	0	126	100	100
42	0	0	-	0	0	113	100	100
43	0	0	-	0	0	141	100	100
44	0	0	-	2	2	123	98.4	98.4
45	0	0	-	0	0	59	100	100
46	0	0	-	0	0	132	100	100
47	0	0	-	3	0	143	97.9	100
48	2	2	100	13	15	97	86.6	84.5
49	0	0	-	0	0	7	100	100
50	0	0	-	1	0	50	98.0	100
51	1	1	100	5	3	66	92.4	95.5
52	0	0	-	0	0	136	100	100
53	1	1	100	2	1	42	95.2	97.6
54	0	0	-	0	0	8	100	100
55	0	0	-	1	0	105	99.0	100
56	0	0	-	10	12	40	75.0	70.0
57	0	0	-	0	0	129	100	100
60	0	0	-	3	1	46	93.5	97.8
61	0	0	-	0	0	11	100	100
62	0	0	-	44	37	108	59.3	65.7
63	0	0	-	2	2	118	98.3	98.3
66	0	0	-	0	0	52	100	100
68	0	0	-	1	1	136	99.3	99.3
Insg.	21	11	52.4	222	168	3594	93.8	95.3

Tabelle C.11: Anzahl alarmrelevanter Zustände bei technisch richtiger Datenaufzeichnung für die Herzfrequenz, Anzahl der vom neuen System erkannten alarmrelevanten Zustände, geschätzte Sensitivität (in %) für den neuen Alarm-Algorithmus, Anzahl falsch positiver Alarme, „Worst-case“-Anzahl und geschätzte Spezifitäten (in %) getrennt für die einzelnen Fälle und insgesamt.

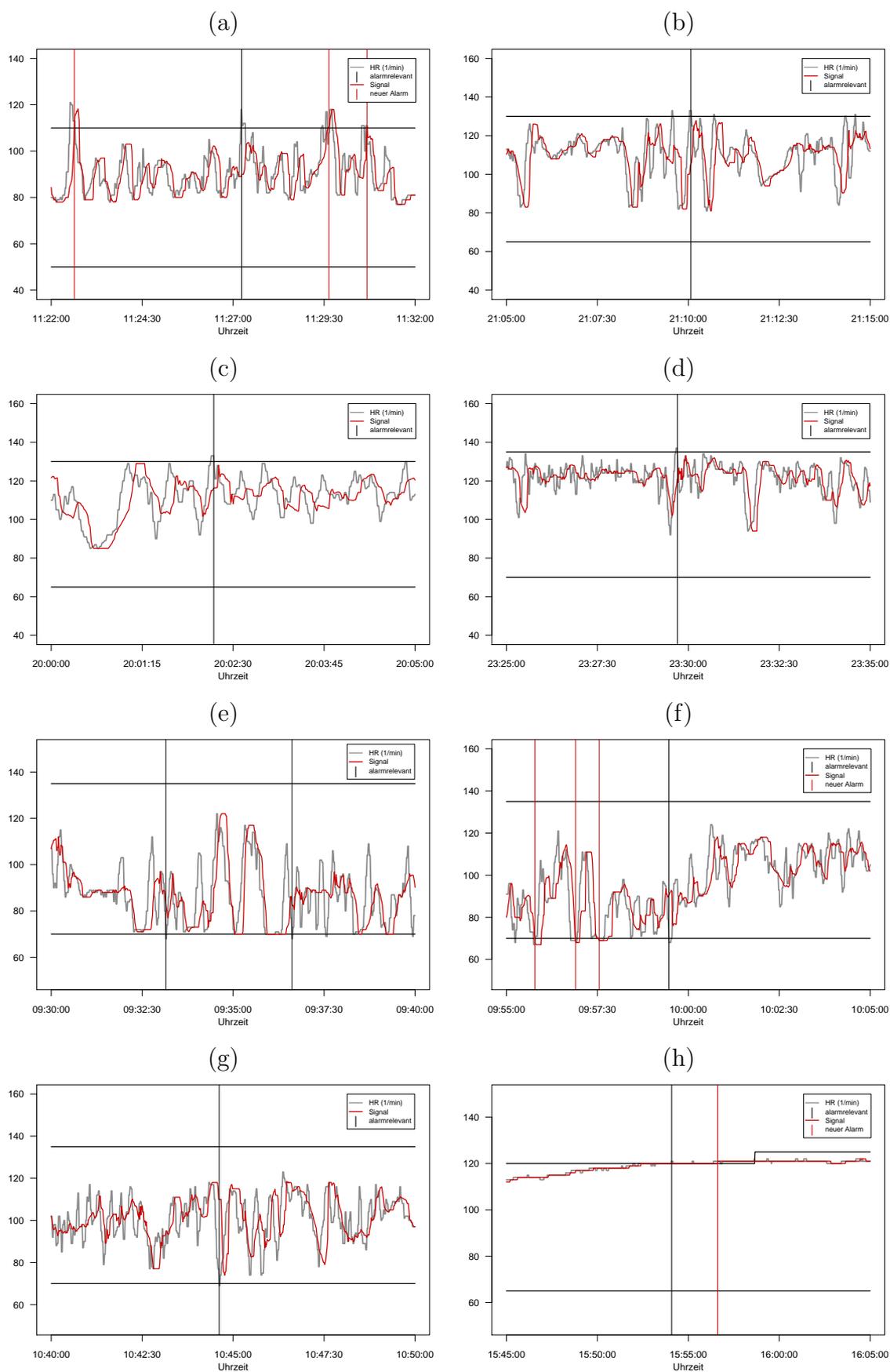


Abbildung C.7: Nicht erkannte alarmrelevante Situationen für die Herzfrequenz

Fall	Anzahl AR (TR)	Anzahl erkannt	$\hat{S}e(A_{neu})$	falsch pos. alt	falsch pos. neu	FP_{max}	$\hat{S}p(A_{alt})$	$\hat{S}p(A_{neu})$
1	1	1	100	3	1	15	80.0	93.3
2	1	1	100	0	0	57	100	100
4	0	0	-	2	2	120	98.3	98.3
5	0	0	-	0	0	4	100	100
8	1	1	100	0	0	113	100	100
9	10	10	100	21	12	49	57.1	75.5
10	19	19	100	14	17	73	80.8	76.7
13	0	0	-	1	1	17	94.1	94.1
14	0	0	-	18	9	57	68.4	84.2
15	0	0	-	0	0	50	100	100
17	0	0	-	2	0	30	93.3	100
18	0	0	-	6	4	100	94.0	96.0
21	0	0	-	6	2	31	80.6	93.5
22	0	0	-	2	3	33	93.9	90.9
23	0	0	-	15	12	62	75.8	80.6
24	1	0	0	0	1	129	100	99.2
26	0	0	-	2	0	92	97.8	100
28	0	0	-	9	9	131	93.1	93.1
29	0	0	-	25	18	131	80.9	86.3
36	3	2	66.7	15	18	81	81.5	77.8
37	2	2	100	15	12	81	81.5	85.2
39	1	1	100	2	4	101	98.0	96.0
40	2	2	100	2	2	98	98.0	98.0
42	0	0	-	1	1	113	99.1	99.1
43	0	0	-	8	6	141	94.3	95.7
44	0	0	-	4	10	123	96.7	91.9
45	2	2	100	7	7	58	87.9	87.9
46	0	0	-	9	10	88	89.8	88.6
47	0	0	-	0	0	143	100	100
48	2	1	50.0	8	6	96	91.7	93.8
49	0	0	-	0	0	7	100	100
50	3	1	33.3	2	1	49	95.9	98.0
51	0	0	-	11	3	66	83.3	95.5
52	0	0	-	2	1	136	98.5	99.3
53	0	0	-	1	2	25	96.0	92.0
54	0	0	-	0	0	8	100	100
55	3	3	100	21	6	40	47.5	85.0
56	0	0	-	1	0	40	97.5	100
57	3	3	100	5	7	128	96.1	94.5
60	0	0	-	0	0	46	100	100
61	1	1	100	1	2	11	90.9	81.8
62	1	1	100	24	18	77	68.8	76.6
63	1	1	100	0	1	118	100	99.2
66	4	3	75	0	0	51	100	100
68	0	0	-	5	8	135	96.3	94.1
Insg.	61	55	90.2	270	216	3354	92.0	93.6

Tabelle C.12: Anzahl alarmrelevanter Zustände bei technisch richtiger Datenaufzeichnung für die Sauerstoffsättigung, Anzahl der vom neuen System erkannten alarmrelevanten Zustände, geschätzte Sensitivität (in %) für den neuen Alarm-Algorithmus, Anzahl falsch positiver Alarmer, „Worst-case“-Anzahl und geschätzte Spezifitäten (in %) getrennt für die einzelnen Fälle und insgesamt.

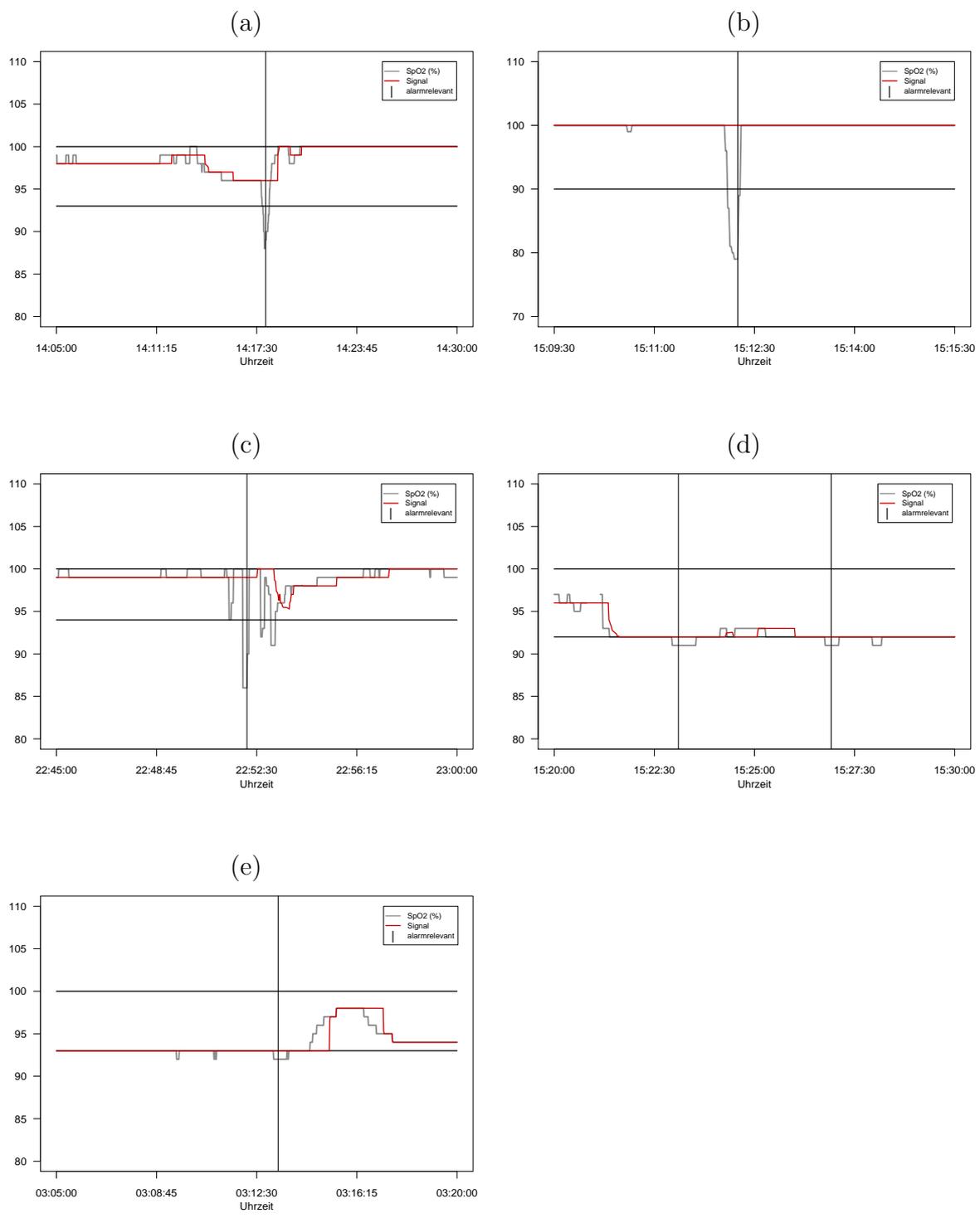


Abbildung C.8: Nicht erkannte alarmrelevante Situationen für die Sauerstoffsättigung

Bezeichnung	Anzahl	Beschreibung						
Gleichzeitige Alarme ($A_{alt}-A_{neu}$)	1218	Dauer (Sek.)	Min. 1	1.Q. 5	Median 13	Mittel 32	3.Q. 35	Max. 364
Unterdrückte Alarme ($A_{alt}-\overline{A_{neu}}$)	259	Dauer (Sek.)	Min. 1	1.Q. 3	Median 5	Mittel 8	3.Q. 9	Max. 201
Alarmrelevant: technisch richtig: 16, technisch falsch: 12								
Gleichzeitige Alarme ($A_{neu}-A_{alt}$)	664	Dauer (Sek.)	Min. 2	1.Q. 36	Median 65.5	Mittel 105.9	3.Q. 131	Max. 1546
		Delay (Sek.)	Min. -196	1.Q. -2	Median 6	Mittel 3.07	3.Q. 18	Max. 56
Zusätzliche Alarme ($A_{neu}-\overline{A_{alt}}$)	152	Dauer (Sek.)	Min. 2	1.Q. 15.75	Median 27	Mittel 35.12	3.Q. 41	Max. 211

Tabelle C.13: Auswertung nach der Methode aus Charbonnier, Gentil (2007) für die Validierungs-Datensätze. Vergleich zwischen alten und neuen Alarmen für die arteriellen Blutdrücke.

Bezeichnung	Anzahl	Beschreibung						
Gleichzeitige Alarme ($A_{alt}-A_{neu}$)	156	Dauer (Sek.)	Min. 1	1.Q. 3	Median 5	Mittel 8.73	3.Q. 10	Max. 107
Unterdrückte Alarme ($A_{alt}-\overline{A_{neu}}$)	90	Dauer (Sek.)	Min. 2	1.Q. 3	Median 3	Mittel 3.67	3.Q. 5	Max. 8
Alarmrelevant: technisch richtig: 9, technisch falsch: 0								
Gleichzeitige Alarme ($A_{neu}-A_{alt}$)	126	Dauer (Sek.)	Min. 1	1.Q. 9	Median 12.5	Mittel 20.93	3.Q. 18	Max. 326
		Delay (Sek.)	Min. -85	1.Q. 0.25	Median 5	Mittel 4.14	3.Q. 9.75	Max. 57
Zusätzliche Alarme ($A_{neu}-\overline{A_{alt}}$)	54	Dauer (Sek.)	Min. 1	1.Q. 6.25	Median 10	Mittel 12.8	3.Q. 13	Max. 135

Tabelle C.14: Auswertung nach der Methode aus Charbonnier, Gentil (2007) für die Validierungs-Datensätze. Vergleich zwischen alten und neuen Alarmen für die Herzfrequenz.

Bezeichnung	Anzahl	Beschreibung						
Gleichzeitige Alarme ($A_{alt}-A_{neu}$)	337	Dauer (Sek.)	Min. 1	1.Q. 9	Median 20	Mittel 42.9	3.Q. 47	Max. 694
Unterdrückte Alarme ($A_{alt}-\overline{A_{neu}}$)	82	Dauer (Sek.)	Min. 1	1.Q. 3.25	Median 5	Mittel 7.23	3.Q. 9	Max. 27
Alarmrelevant: technisch richtig: 6, technisch falsch: 6								
Gleichzeitige Alarme ($A_{neu}-A_{alt}$)	222	Dauer (Sek.)	Min. 2	1.Q. 47	Median 88.5	Mittel 154.9	3.Q. 194.5	Max. 1261
		Delay (Sek.)	Min. -196	1.Q. -6	Median 3	Mittel 2.667	3.Q. 16	Max. 386
Zusätzliche Alarme ($A_{neu}-\overline{A_{alt}}$)	81	Dauer (Sek.)	Min. 1	1.Q. 26	Median 70	Mittel 118.8	3.Q. 149	Max. 1906

Tabelle C.15: Auswertung nach der Methode aus Charbonnier, Gentil (2007) für die Validierungs-Datensätze. Vergleich zwischen alten und neuen Alarmen für die Sauerstoffsättigung.