needed. Nearly all clinical studies into monitoring alarms used clinician judgement and annotation as the reference standard. We investigated the intra-observer and inter-observer variability between two intensivists in the classification of monitoring time series.

Methods A total of 3,092 time series segments (heart rate and blood pressures) of 30 minutes each from six critically ill patients were presented to two experienced intensivists (MD1 and MD2) offline and were visually classified into clinically relevant patterns (no change, level shift, trend) by the physicians separately. One intensivist (MD2) repeated the classification 4 weeks after the first analysis on the same dataset.

Results MD1 found clinically relevant events in 36%, and MD2 in 29% of all time series. In 16% of all cases both intensivists came to different classifications. In 10% even the direction of change was classified differently. MD2 classified 10% of all cases differently between the first and second analysis. Even if level changes and trends were treated as one universal pattern of change, intra-individual variability (MD2 first analysis vs MD2 second analysis) was still 5% and inter-individual variability (MD1 vs MD2, only unequivocal classifications) was 10%.

Conclusion Although this study is small with only two observers who were investigated, it clearly shows that there is a significant intra-individual and inter-individual variability in the classification of monitoring events done by experienced clinicians. These findings are supported by studies into image analysis that also found high intra-individual and inter-individual variability. High inter-observer and intra-observer variability is a challenge for clinical studies into new alarm algorithms. Our findings also show a need for reliable classification methods.

P438

Robust regression methods for intensive care monitoring

M Imhoff[1], K Schettlinger[2], R Fried[2], U Gather[2], S Siebig[3], C Wrede[3]

[1]Ruhr-University Bochum, Germany; [2]University of Dortmund, Germany; [3]University Hospital Regensburg, Germany

Critical Care 2007, **11(Suppl 2):**P438 (doi: 10.1186/cc5598)

Introduction Alarm generation of modern patient monitoring systems still predominantly relies on simple threshold methods. This leads to an unacceptably high rate of false positive alarms. Many false positive alarms are generated by measurement artefacts and measurement noise. One approach to address this problem is to alarm on the underlying signal (that is, the noise-free time series of the physiological variable), instead of the raw measurement.

Methods Monitoring time series were simulated. Against these data four robust regression methods were evaluated: least trimmed squares (LTS), least median of squares (LMS), repeated median (RM), and deepest regression (DR). Moreover, online monitoring series from critically ill patients during multiparameter monitoring were also compared.

Results LTS and LMS showed comparable behaviour, as did RM and DR. LMS and LTS provided only 20% efficiency, DR 61% and RM 70% (least squares regression = 100%). RM and DR had smaller standard deviations and smaller mean-squared errors than LMS and LTS under different noise distributions (standard deviation of online estimates based on sliding windows of size $n = 21$ for simulated standard normal errors: LMS: 0.875, LTS: 0.887, RM: 0.500, DR: 0.533). Analyses with clinical monitoring data also showed that LMS and LTS preserve sudden level shifts but are unstable and perform poorly with trend changes; RM and DR blur shifts but yield more stable estimations.

Conclusion All four methods allow one to extract the underlying signal from physiological time series in a way that is robust against measurement artefacts and noise. However, there are significant differences between the methods. Overall, repeated median regression seems the best choice for intensive care monitoring since it is not only the most stable but also the fastest method.

P439

ISIS program: a new tool for medical research at the bedside in critical care units

H Mehdaoui[1], B Sarrazin[1], I El Zein[1], L Allart[2], C Vilhelm[2], S Guerra[2], D Zitouni[2], M Lemdani[2], R Valentino[1], A Herbland[1], P Ravaux[2]

[1]Fort De France University Hospital, Fort De France, Martinique; [2]Lille 2 University, Lille, France

Critical Care 2007, **11(Suppl 2):**P439 (doi: 10.1186/cc5599)

Introduction The goal of this program is to develop an experimental tool able to record, store and analyse data issued from critical care patients. Due to technical limitations and medical constraints, information systems able to manage such data flow are difficult to deploy.

Methods Data recording is done through a laptop connected to the medical devices, allowing analogical and digital signal transmission through a high-speed network. Several servers are dedicated to specialised tasks: mass storage, model generation, artificial intelligence (AI), telecommunications, and security. A 3 Teraflops supercomputer is dedicated to intensive computation when necessary. Twenty applications are dedicated to elective tasks, most of them running using the Linux operating system.

The 'Aiddiag' data-acquisition software is a standalone application adapted to patient data recording from the biomedical devices and caregiver's inputs. It has a friendly designed user-interface touchscreen at the bedside and was adapted according to caregivers' feedback. Data are also stored in a repository and a selective secondary extraction is possible. Online and offline analysis by the AI engine is allowed. Software had to consider time specifications and uses distributed computation to achieve high workload tasks. We complied to the French legal patient data management constraints.

Results After 2 years, our system is fully deployed. It recorded more than 2,500 patient-hours over a 3-month period. Signal loss is less than 1%. Our tool allows recording of more than 40 digital signals, eight analogical signals sampled at a rate of 1 kHz, and caregiver comments and actions. CPU resources of the laptop are available for supplemental AI developments during data acquisition. Transfer of data to the repository is either a hotplug-automated process or delayed with 5 days of buffering in the laptop. Automated artefacts' cleaning allows time-series analysis (GARCH method) to extract behavioural models after intensive computation. The AI engine is used for medical guideline implementation (that is, severe brain trauma care algorithms) and later comparison with caregiver's behaviour. Remote use of our system is possible and schedulable, allowing other research teams to work on the data. Limitations have been detected during intensive calculation. Fine-tuning of the network will suppress these limitations.

Conclusion ISIS is the first program to achieve an easy-to-use recording tool able to build a very large medical repository. Data analysis methods and AI-controlled automated complex medical guidelines are under evaluation.