©Digital Vision

The Challenges in Creating Critical-Care Databases

Dealing with the Special Requirements of Obtaining Data in a Demanding Environment

Ilkka Korhonen[1], Mark van Gils[1], John Gade[2]

[1] VTT Information Technology, Tampere

[2] Judex Datasystemer, Aalborg

Intensive care is required when the patient's vital functions are disturbed due to a disease or trauma, and their maintenance requires continuous external support to allow time for restoration and normalization of these functions. In general anesthesia induced to provide necessary hypnosis, analgesia, and muscle relaxation to tolerate painful stimuli caused by surgery and to enable safe operation, the patient's vital functions are similarly dependent on life-supporting devices and intensive monitoring. In these critical-care set-ups, a large variety of technical devices are available and produce a continuous flow of data, which need to be integrated with other data, such as laboratory test results, background information, previous diagnosis, care plan, medication, etc. Hence, there is a risk for "information overload," an inability of the care staff to efficiently use all the information available about the patient state. More intelligent and advanced patient monitoring may reduce this overload [1].

As critical care is critical, the demands for the patient monitoring are high. The methods need to function properly over a large variety of possible physiological conditions in an environment full of sources for technical artifacts, and they need to support clinical reasoning. Several iterations are usually required in the method development to meet these demands [2]. First, the proposed method is developed and tested with simulated or optimal test data on a proof-of-principle basis. When this level is passed, the method needs to be tested offline with more realistic and nonoptimal data. Only after passing these levels may clinical trials be conducted. Usually, these developments require a significant period of time. Availability of extensive, well-characterized, and well-documented real physiological data during critical care would potentially reduce the amount of time required for the first phases of the method development [2]. Such databases would also enable bench validation of any new methods and objective comparison between different methods.

Recently, there have been some attempts to collect large signal databases during critical care. In this article, we aim to discuss the special requirements for collecting signal databases in critical care and to summarize the main features of the existing databases.

Collecting Databases During Critical Care

Collecting an extensive and well-characterized physiological database during critical care is complicated for various reasons:

1) *On-going care*. The most important difficulty is the fact that little modifications are allowed for routine patient care in intensive care or surgery due to data collection projects. This means that the routine patient monitoring and support devices need to be attached as usually and required therapy will be given without compromises and in time despite the on-going data collection. Hence, there will be little possibility to employ complicated standardized study protocols, to collect artifact-free reference data, to use time-consuming specific stimulation to test the responsiveness of the patient, or even to attach extra sensors or electrodes to the patients.

2) *Recording devices*. Technical properties of the devices in clinical use do not always satisfy the scientific requirements, but yet they may be impossible to replace due to lack of space and the demand to guarantee routine patient care also in study patients. For example, if a routine Swan-Ganz catheter is used, it is impossible to use another model with the same patient simulta-

neously. Hence, in practice one usually needs to employ a standard patient monitor with its built-in features in the collection of most of the signals. Thus, compromises are needed in issues such as electrode or sensor types and placements, sampling rates, and level of signal conditioning.

3) *Resources and delays*. Especially in intensive care it is very difficult to predict the timing of the patient admission, even more so if patients with certain pathologies are to be considered. Without a significant overhead it is impossible to guarantee the availability of the necessary research staff and equipment exactly when needed. Furthermore, the need for consent for the study required in advance from the patient or his/her relatives leads to an unavoidable delay from the admission to the study start. This is unfortunate since, often, the time just after the admission represents the most unstable period in the patient state and would be of special interest.

These factors have the result that the data collection in critical care will by necessity be typically more a comprehensive documentation of the on-going care than a well-controlled clinical study inducted through a well-defined study protocol. This is probably the most important difference in dealing with critical care to most other types of biomedical signal database collection projects.

Existing Databases

To our knowledge there exist four critical-care databases available or to be available to the public. In the following, these databases are briefly summarized. A comparison of their most important features is provided in Table 1.

MGH/MF Waveform Database

Massachusetts General Hospital/Marquette Foundation (MGH/MF) Waveform Database1 was collected in Massachusetts General Hospital critical-care units, during surgery, cardiac catheterization, or other electrophysiology studies in 1992, and it is the oldest of the critical-care databases. It is a collection of hemodynamic and electrocardiographic signals with beat labels, event annotations, and relevant clinical data. The event annotations include, for example, offline labeling of single heartbeats. As its name states, it is a waveform database and does not have processed trends or other supporting data available. The MGH/MF Waveform database is available for licensing from Massachusetts General Hospital (Anaesthesia/Bioengineering Unit, Massachusetts General Hospital, Boston, MA, USA).

MIMIC

The collection of the MIMIC (Multi-parameter Intelligent Monitoring for Intensive Care) Database was started in 1994, and still continues. Eventually, it will include 100 patient records containing both the most important continuously monitored waveforms and the patient's clinical data [3]. The 24- to 48-hour-long records from patients in medical, surgical, and cardiac intensive care units of Beth Israel Hospital, Boston, MA, USA, are collected by an automatic procedure using Hewlett-Packard Component Monitoring System (Merlin) patient monitors. The patients with high probability for unstable hemodynamics are recruited. The data will include detailed clinical data from a patient's medical record and hospital online medical information systems. A sample record of MIMIC is available through PhysioNet [4], but the price or availability of the full database has not been announced as yet.

IMPROVE

In the IMPROVE (Improving Control of Patient Status in Critical Care) project, an annotated data library was constructed from recordings during intensive care at Kuopio University Hospital, Kuopio, Finland [5]. The recordings were made through a standard Datex AS/3 patient monitor and related clinical systems. However, for seven cases an extra ADAM (Advanced Depth of Anesthesia Monitor) system was used to record the electroencephalogram (EEG) [6]. The annotation of the data library meant that the patient state and clinical interventions were continuously monitored and recorded by a member of a team of physicians who remained at the bedside throughout the recording, which typically lasted 24 hours. The data recorded included practically all routinely monitored continuous waveforms, together with all other clinical data obtained during the routine course of care. By annotating the recording during data collection it was ensured that any changes in the recorded waveforms or measurements could be referred against the physician's observations. The patients included were those with a high risk of oxygen-transport-related problems during the recording. The IMPROVE Data Library has been available since 1997 through VTT Information Technology, Tampere, Finland.

Data collection in critical care will by necessity be typically more a comprehensive documentation of the on-going care than a well-controlled clinical study.

IBIS

The IBIS (Improved monitoring for Brain dysfunction in Intensive care and Surgery) data library was collected between 1998 and 1999 by largely the same team as the IMPROVE data library [7-9]. It has been collected in three hospitals: one in Finland (intensive care unit of Kuopio University Hospital) and two in London, UK (operation theaters in Saint Bartholomew's Hospital, and operation theaters and intensive care unit of Royal Brompton Hospital); hence, it differs from the above-mentioned databases in the sense that it is a multicentric database. The IBIS data library concentrates on neuromonitoring (i.e., the well being of the central nervous system) during critical care. This is achieved by monitoring EEG, evoked potentials (EP), and event-related potentials (ERP) together with electrocardiogram (ECG) and other routinely monitored data. As with the IMPROVE data library, all routinely monitored data is acquired by using commercial patient monitors and information management systems from Datex-Ohmeda Ltd. as well as a modified infra-operative monitoring system from Nicolet Biomedical Inc. The recordings are obtained during restoration from cardiac surgery; during cardiac, vascular or neurosurgery; or from multiple organ failure patients in intensive care; all with a high risk for central nervous system

Table 1. Comparison of Different Critical Care Databases				
	MIMIC	**MGH/MF waveform**	**IMPROVE**	**IBIS**
type of data	surgical and cardiac ICUs	critical care units, during surgery, cardiac catheterization or other electrophysiologic studies	intensive care patients with defined clinical disorders	ICU and OR patients emphasis on neuromonitoring
categories	hemodynamically unstable patients	ventricular arrhythmias atrial arrhythmias CO2 and impedance rec	hypovolaemia, cardiac failure, high-blood flow state and O_2 related problems	cardiac failure, MOF (ICU) anaesthesia during operations (cardiac, vascular, neuro) (OR)
cont signals/waveforms				
hemodynamics	SAP, PAP, plethysmography	SAP, PAP, CVP (intra-cranial, left atrial, ventricular and/or intra-aortic balloon pressures included in some cases)	SAP, PAP, CVP	SAP, PAP, CVP
ECG	2-3 channels	3 channels	2 channels	2 channels
respiratory signals	AWP, CO_2	impedance, airway CO_2	AWF, AWP, O_2, CO_2	AWF, AWP, O_2, CO_2, AA
EEG	no	no	2 channels (7 cases only)	2 channels (5 channels intermittently)
evoked potentials	no	no	no	EP: MLAEP, BAEP, SEP ERP: MMN, HabN100
trend/intermittent data				
hemodynamics/ respiratory signals	all monitored variables are included at intervals of 1.024 s	None	all monitored hemodynamic and respiration-related trends (1 value / 2 mins)	all monitored hemodynamic and respiration-related trends (1-2 values / 2 mins)
other			Temperatures + all routinely measured data in ICU	Temperatures + all routinely measured data in ICU or OR
laboratory tests	yes		yes	yes
annotations	patient status alarms / monitor alarms (by monitor)	Beat labels and relevant events	patient state, nursing actions, disturbances/artifacts (by physician)	patient state, free text (by physician)
size				
number of recordings	100	250	59	48 (ICU) + 52 (OR)
number of patients	100	225 adults + 25 pediatric	50	100
recording sites	ICUs	ICU, catheritization lab and OR	ICU	2 ICUs, 2 ORs
length of typical recording	24-48 hrs	90 mins	24 hrs	3 hrs
total length of data	Not available	375 hrs	>1300 hrs	ICU: 171 hrs, 241 ERPs, 1304 EPs OR: 173 hrs, 281 ERPs, 2181 EPs
distribution				
main data format	MIT-BIH	MIT-BIH	EDF & ASCII	EDF, XEP & ASCII
medium	WWW download & CD-ROMs	10 CD-ROMs	9 CD-ROMs	>12 CD-ROMs
case documentation	yes	yes	yes	yes
tools available	PhysioToolkit (GNU GPL)	PhysioToolkit (GNU GPL)	Win9x/NT4/3.x tools included other EDF browsers supported Matlab® routines	Win9x/NT4/3.x tools included other EDF browsers supported Matlab® routines
availability/pricing	to be announced	12.000 USD	4.000 Euro (research) / 10.000 Euro (industry)	to be announced
web page	www.physionet.org/ physiobank/database/mimicdb/	ecg.mit.edu/ dbinfo.html#DB Software Package	www.vtt.fi/tte/samba/projects/improve/improve.htm	www.vtt.fi/tte/samba/projects/ibis/data/data.html
notes	not yet available (partly via WWW site)			not yet available

Practical problems include resource management, patient consent queries, technical problems due to the high load of the systems and lack of space, etc.

hypoxia. The annotations resemble those of the IMPROVE data library although concentrating more on the neuromonitoring domain and being less intensive; e.g., for artifacts.

The IBIS data library will be made available for third parties during 2001, and will be delivered like the IMPROVE data library through VTT Information Technology.

Discussion

Collecting a large, well-characterized database during critical care is a laborious and demanding but also very rewarding task. The practical problems include resource management, patient consent queries, technical problems due to the high load of the systems and lack of space, limited possibilities to modify either the monitoring setup or care to allow specific study protocols and settings, etc. This sets boundaries to the data collection, which becomes an extensive documentation of a certain period of care rather than a standardized, well-controlled study. These conditions are important to keep in mind while planning to create or using any critical care database.

Clinical decision making is critically dependent on the ability to record and analyze physiological signals [4]. As this dependency is critical, the adoption of any new signal processing method into clinical practice in critical care is slow [2], and relatively conventional methods are still the state-of-art in clinics despite the technological advances in signal processing resources and methods [4]. As Goldberger and co-workers point out, this adoption is slowed down by the lack of three types of resources [4]:

1) *Data resources* (i.e. researchers developing and testing new methods lack access to necessary data, both in terms of quality, quantity, and representativeness);

2) Analytic resources (i.e., resources for new and creative analysis run short while significant resources in research projects are spent on "re-inventing the wheel": implementing analysis tools to get access to data, and to run through basic preprocessing and analysis); and

3) Human and communication resources (i.e., development of complex biomedical signal analysis requires simultaneous expertise from many fields, and this is rarely available within a single research group).

In critical care these points are indeed emphasized, as both the demands for the methods and the complexity of the environment are very high. The existing databases for critical care aim to relieve the lack of especially the first of the resources mentioned above. As large, documented and well-characterized sources of real data, they allow a short cut for method development and testing as well as objective comparison of performance of different methods. The data are provided in supported data formats, such as European Data Format [10], and tools to access the data are readily available with the databases; this should relieve the lack of analytic resources, (point 2 above). Finally, these large efforts of collecting and delivering the data encourages building up international "user clubs," which potentially exceed the "critical mass of expertise" referred to in point 3 above.

The existing critical care databases all provide large amounts of real data in critical care yet differ in many details and characteristics (Table 1). These differences have some implications for their suitability for different purposes. The annotations in the IMPROVE and IBIS data libraries especially differ from those in MGH/MF Waveform and MIMIC databases. In IMPROVE and IBIS, a physician observed the patient state and monitored signals online and annotated any changes in patient state or possible external causes for artifacts. This is a very labor-intensive task but ensures capture of many such pieces of information that would be missing were the annotations made offline afterwards. Another difference in annotations between these databases is that in IMPROVE and IBIS the annotations do not extend into such time resolution as with, for example, the MGH/MF Waveform database where single heart beats are labeled. These differences make the IMPROVE and IBIS data libraries more suited for development of algorithms to predict overall patient status and outcome, while, for example, for development of an optimal ECG analyzer for critical care the MGH/MF Waveform database would be a better choice.

The MGH/MF Waveform database and IMPROVE data library have now been available for some years, while MIMIC

Table 2. Abbreviations Used in Table 1

Abbreviation	Meaning
AA	anesthetic agent
AWF	airway flow
AWP	airway pressure
BAEP	brainstem auditory evoked potential
CO_2	carbon dioxide
CVP	central venous pressure
ECG	electrocardiogram
EDF	European data format
EEG	electroencephalogram
EP	evoked potential
ERP	event related potential
HabN100	habituation of N100 (protocol to study ERPs)
ICU	intensive care unit
MIT-BIH	Massachusetts Institute of Technology - Beth Israel Hospital data format
MLAEP	mid-latency auditory evoked potential
MMN	mismatch negativity (protocol to study ERPs)
O_2	oxygen
OR	operation room
PAP	pulmonary arterial pressure
SAP	systemic arterial pressure
SEP	somatosensory evoked potential

and IBIS databases are yet to be released. So far, the MEDLINE database reports no studies based on the MGH/MF Waveform database. In turn, the IMPROVE [11] and IBIS [12] projects have both published a special issue reporting the first results based on the usage of the corresponding databases. However, the final success of these databases may only be judged after some years when we have seen the true extent of their usage and may evaluate the results achieved by using them.

Ilkka Korhonen was born in 1968 in Hankasalmi, Finland. He received his M.Sc. and Dr.Tech. degrees in digital signal processing from Tampere University of Technology in 1991 and 1998, respectively. He is currently working at VTT Information Technology. He is a docent in medical informatics (with speciality in biosignal processing) at the Ragnar Granit Institute at Tampere University of Technology and a member of IEEE EMB Society. His main research interests are to apply biosignal interpretation methods in critical-care patient monitoring, autonomous nervous system research, and home or remote health monitoring.

Mark van Gils received his M.Sc. degree in applied physics from the Technical University of Eindhoven in 1990. In 1995, he obtained a Ph.D. degree from the same university at the Faculty of Electrical Engineering for research on the application of artificial neural networks techniques on feature extraction from evoked potential recordings. After that he worked as a visiting scientist at the NIH/NIA Gerontology Research Center in Baltimore, MD, USA in 1995. He is currently employed as a senior research scientist at VTT Information Technology and holds an appointment as docent in the area of physiological signal processing at the Technical University of Helsinki. His research interests include, but are not limited to, physiological signal processing, information visualization, (bio)signal interpretation, as well as technology transfer.

John Gade received his M.Sc. degree in electronics engineering from Aalborg University, Denmark, in 1988. He worked as a research assistant and assistant professor in the Department of Medical Informatics and Image Analysis of Aalborg University in 1988-1995. Currently, he is employed as a product manager at Judex A/S, Aalborg, Denmark, as well as a lecturer in health informatics at Aalborg University. His R&D interests include processing, analyzing, and visualizing information present in biomedical signals.

Address for Correspondence: Ilkka Korhonen, VTT Information Technology, P.O. Box 1206, FIN-33101 Tampere, Finland. Tel: +358 3 316 3352, Fax: +358 3 317 4102. E-mail: ilkka.korhonen@vtt.fi.

References

[1] T. Sukuvaara and E.M.J. Koski, "Informative alarms in anaesthesia: from signal to patient-state monitoring," *Current Opinion in Anaesth*.,vol. 8, no. 6, pp. 526-531, 1995.

[2] N. Saranummi, I. Korhonen, M. van Gils, and A. Kari, "Framework for biosignal interpretation in intensive care and anaesthesia," *Methods Inform. Medicine*, vol. 36, pp.340-344, 1997.

[3] G.B. Moody and R.G. Mark, "A database to support development and evaluation of intelligent intensive care monitoring," *IEEE Comp Cardiol*., pp. 675-660, 1996.

[4] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, PCh. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C-K Peng, and H.E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet. Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215-e220, 2000.

[5] I. Korhonen, J. Ojaniemi, K. Nieminen, M. van Gils, A. Heikelä, and A. Kari, "Building the IMPROVE Data Library," *IEEE Eng. Med. Biol. Mag.,* vol. 16, no. 6, pp. 25-32, 1997.

[6] C.E. Thomsen, J. Gade, K. Nieminen, R.M. Langford, I.R. Ghosh, K. Jensen, M. van Gils, A. Rosenfalck, P. Prior, and S. White, "Collecting EEG signals in the IMPROVE Data Library. Data acquisition and visual analysis tools for obtaining prolonged recordings in intensive care," *IEEE Eng. Med. Biol. Mag*., vol. 16, no. 6, pp. 33-40, 1997.

[7] J. Gade, I. Korhonen, M. van Gils, P. Weller, and L. Pesu, "Technical description of the IBIS Data Library," *Comp. Meth Prog Biomed*., vol. 63, no. 3, pp. 175-186, 2000.

[8] S.M. Jakob, K. Nieminen, J. Karhu, and J. Takala, " IBIS data library: clinical description of the Finnish database," *Comp Meth Prog Biomed.,* vol. 63, no. 3. pp. 161-166, 2000.

[9] G.F. Mandersloot, R.C. Pottinger, P.R. Weller, P.F. Prior, C. Morgan, N.J. Smith, and R.M. Langford, "The IBIS Project. Data Collection in London," *Comp Meth Prog Biomed.*, vol. 63, no. 3, pp. 167-174, 2000.

[10] B. Kemp, A. Värri, A.C. Rosa, K.D. Nielsen, and J. Gade, "A simple format for exchange of digitized polygraphic recordings," *Electroencephalogr. Clin. Neurophysiol.*, vol. 82, pp. 391-393, 1992.

[11] The IMPROVE Project. Special Issue of *IEEE Eng. Med. Biol. Mag.*, vol. 16, no. 6, 1997.

[12] The IBIS Project. Special Issue of *Comp Meth Prog Biomed*., vol. 63, no. 3, 2000.