

Original Paper

Feasibility of a German High-Frequency Neonatal and Pediatric Intensive Care Dataset (Alx-Neo-Guard Dataset): Secondary Data Cohort Study

Lena Sophie Olivier¹, Dr med; Camelia Lauterbach Oprea², MSc; André Stollenwerk², Dr-Ing; Thorsten Orlikowsky¹, Prof Dr med; Mark Schoberer¹, Priv-Doz Dr med

¹Division of Neonatology, Department of Pediatric and Adolescent Medicine, Uniklinik RWTH Aachen, Aachen, Germany

²Chair of Embedded Software (Computer Science 11), RWTH Aachen University, Aachen, Germany

Corresponding Author:

Lena Sophie Olivier, Dr med

Division of Neonatology, Department of Pediatric and Adolescent Medicine

Uniklinik RWTH Aachen

Pauwelsstraße 30

Aachen, 52074

Germany

Phone: 49 2418089230

Email: Lolivier@ukaachen.de

Abstract

Background: Available information about existing neonatal and pediatric intensive care datasets is scarce.

Objective: The objective was to evaluate the feasibility of a monocentric high-frequency dataset of neonatal and pediatric intensive care patients in a cohort study.

Methods: This study included patients treated in the neonatal and pediatric tertiary care intensive care unit at Uniklinik RWTH Aachen, Germany. The dataset comprised high-frequency data on bedside monitoring and ventilation, CO₂ monitoring, laboratory results, important diagnoses and interventions, patient characteristics, and manual annotations (including procedures, eg, intubation and X-rays; and states and illnesses, such as episodes of apnea of prematurity and patient-ventilator asynchronies).

Results: In total, 400 admissions (227 neonatal and 173 pediatric) from 359 patients were included between September 2024 and January 2026. Of these, 28 neonates were extremely preterm, 34 were very preterm, 68 were moderate to late preterm, and 97 were term infants, with an overall median gestational age of 35 (IQR 31-38) weeks and a median birth weight of 1805 (IQR 1446-3251) g. The pediatric population comprised 34 infants aged younger than 1 year, 45 toddlers aged 1 to 5 years, 43 children aged 6 to 12 years, and 51 adolescents aged 13 to 18 years, with a median age of 10.6 (IQR 1.2-13.0) years and a median admission weight of 22 (IQR 11-50) kg. Overall, 77% (400/517) of all eligible patients were enrolled in the study. The dataset comprised 3600 patient days with more than 20,000 manual annotations (approximately 3 TB) and has been used for multiple use cases.

Conclusions: The high temporal resolution allows for detailed characterization of patient states. To our knowledge, there are no published descriptions of comparable European high-frequency neonatal or pediatric intensive care datasets. We share information about our dataset to encourage cross-institutional cooperation. The multicenter pooling of data or federated learning increases possible use cases by enabling new research questions or the development of more robust algorithms.

(*JMIR Pediatr Parent* 2026;9:e94280) doi: [10.2196/94280](https://doi.org/10.2196/94280)

KEYWORDS

critical care; database; pediatrics; neonatology; pediatric intensive care units; neonatal intensive care units; mechanical ventilation; data warehousing

Introduction

Approximately 12% of neonates and 13% of hospitalized children require treatment in an intensive care unit (ICU) [1,2].

Although there have been advances in neonatal and pediatric research in the last decades, many common treatment strategies are still not evidence based [3]. Scientific findings from studies in adult populations are rarely validated on pediatric data [4].

Furthermore, they often cannot be transferred to pediatric patients due to substantial differences in pathophysiology [5]. Despite frequently unknown risks, this results in the widespread off-label use of medication, accounting for up to 50% of prescriptions [6]. This gap of knowledge is particularly problematic because childhood is a critical period for disease development, with long-lasting implications on personal health [7,8].

Neonatal and pediatric intensive care research is challenged by a heterogeneous patient population (eg, with a wide weight range) and small sample sizes [9-11]. Ethical considerations, legal frameworks, and parental uncertainty often obstruct study participation [12]. It can be conflicting to find a balance between protecting the most vulnerable patients by avoiding stress associated with clinical trials on the one hand and obtaining scientific evidence for adequate therapies on the other hand [13]. This leads to recruitment challenges with limited funding and persistently low study enrollment of less than 50% of eligible patients [14-16].

Secondary data research refers to the use of preexisting data that were collected for other purposes and are reused for scientific investigations [17]. It is time-, cost-, and resource-efficient and often faces fewer bureaucratic obstacles than experimental studies [18]. Among others, data can be used for complication detection, clinical decision support, quality analysis, and individualized medicine [19]. It can be helpful in gaining a deeper insight into the effects and side effects of off-label treatments [20]. Despite their advantages and the opportunity to address a variety of unanswered research questions, secondary data are still underused [21].

Data storage solutions for high-frequency intensive care data (such as the Philips Data Warehouse Connect) have been commercially available for almost a decade and are increasingly used. It can be assumed that this is also true for pediatric and neonatal intensive care data, although only very few such datasets are openly available. The storage capacity of these systems exceeds that of *conventional* patient data management systems (PDMSs) by several orders of magnitude. It allows for storing not only numerical but also waveform data at sampling frequencies up to 500 Hz. In turn, this enables researchers to reconstruct critical clinical situations at a high temporal resolution. We established a Data Warehouse Connect system and conducted a study of our pediatric and neonatal patient cohort beginning in August 2024. With regard to data composition, pediatric and neonatal intensive care treatment is characterized by above-average lengths of stay and relatively low patient counts [22]. This poses a dilemma for data research: the limited patient numbers demand cross-institutional data pooling for many research questions [23]. At the same time, low patient counts make it particularly difficult to comply with data protection requirements for anonymized data such as *k*-anonymity [24]. Our institutionally approved data protection concept therefore demands only sharing anonymized data with third parties for defined purposes with a restricted user circle under contractual protection. The execution recommendations of the Technology and Methods Platform for Networked Medical Research e.V. also advise against the disclosure of anonymized data in public-use files unless “based on general

life experience or the current state of scientific knowledge, it is practically impossible to link the data to a specific individual” [25]. The primary objective is to publish the dataset. We hope that, as we move closer to concepts such as the European Health Data Space—including the needed underlying legal perspective—we will be in a position to make the data available to the community in the future. Until then, we see collaborative research as a viable option. This approach allows AI algorithms to be trained independently at different locations using locally available data, such that only model parameters are shared and no explicit patient data are exchanged; this approach is known as federated learning [26]. Another option is to perform external validation for the developed algorithms.

We present our project and metadata with two objectives: (1) to enable potential collaboration partners in our field to determine the extent to which our data and research activities align with their own and (2) to demonstrate that meaningful secondary data analysis can be conducted under the challenging conditions of contemporary neonatal research, even in single centers, and to motivate others to engage in this field [27].

Methods

Study Design

Aix-Neo-Guard is a collaborative project of the neonatal and pediatric ICU of Uniklinik RWTH Aachen, Germany, and the Chair of Embedded Software of RWTH Aachen, Germany. The project is conducted as an observational cohort study and was prospectively registered at the German Clinical Trials register (DRKS00033228). The study is conducted in accordance with the Declaration of Helsinki [28] and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement guidelines for reporting observational studies [29]. Data collection started in September 2024 and has no scheduled end date.

Ethical Considerations

The study was approved by the local ethics committee (Ethik-Kommission an der Medizinischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen [RWTH Aachen]; approval EK 24-213). The German Health Data Utilization Act grants an exemption from individual parental informed consent, as only data that are regularly used for diagnostic and therapeutic purposes are collected. Openly available information about ongoing studies at an institution needs to be provided to enable the exemption, and parents can actively revoke study participation.

Setting

Data are collected at the neonatal and pediatric tertiary care ICU of Uniklinik RWTH Aachen, Germany. The ward offers 9 beds, each equipped with the capability to provide respiratory support. Annually, approximately 350 patients aged 0 to 17 years are treated on the ward.

Participants

All eligible patients treated in the ICU are included in the dataset. Patients are excluded if their parents object to data

processing or study participation. Furthermore, technical issues can lead to unintended exclusion.

Variables

The dataset incorporates monitoring and ventilation data, laboratory results, patient characteristics, selected diagnoses

and interventions, and manual annotations. An overview is given in [Table 1](#). Waveform data are stored at 125 Hz, except for electrocardiogram (ECG) data, which are stored at 500 Hz. Numerical data are recorded with a sampling rate of 1 Hz.

Table 1. Available parameters with their temporal resolution and source.

Category and examples	Availability or resolution	Source
Vital signs		Bedside monitoring
Heart rate	1 Hz	
Respiratory rate	1 Hz	
ECG ^a	500 Hz	
Chest impedance	125 Hz	
Body temperature	1 Hz	
Arterial waveform and pressure	125 Hz	
Noninvasive blood pressure	Intermittent	
Ventilation data		Ventilator
Settings		
FiO ₂ ^b	1 Hz	
PEEP ^c	1 Hz	
Waveforms		
Airway pressure	125 Hz	
Airway flow	125 Hz	
Measurements		
Measured PEEP	1 Hz	
Tidal volume	1 Hz	
CO ₂ monitoring		Bedside monitoring
End-tidal	62.5 Hz	
Transcutaneous	1 Hz	
Laboratory results	Intermittent	PDMS ^d
Blood gas analyses		
Blood count		
Serum chemistry		
Urine analyses		
Patient characteristics	One time	PDMS
Sex		
(Gestational) age		
Admission weight		
Umbilical cord pH		
Maternal age		
Manual annotations	Partially	PDMS and manual data analyses
Apnea of prematurity		
Times of medical imaging		
Nursing		
Important interventions		
Intubation		
Time of death		
Patient-ventilator asynchrony		
Body positioning	Partially	Paper documentation

Category and examples	Availability or resolution	Source
Prone position		
Kangaroo care		

^aECG: electrocardiogram.
^bFiO₂: fraction of inspired oxygen.
^cPEEP: positive end-expiratory pressure.
^dPDMS: patient data management system.

Data Sources and Management

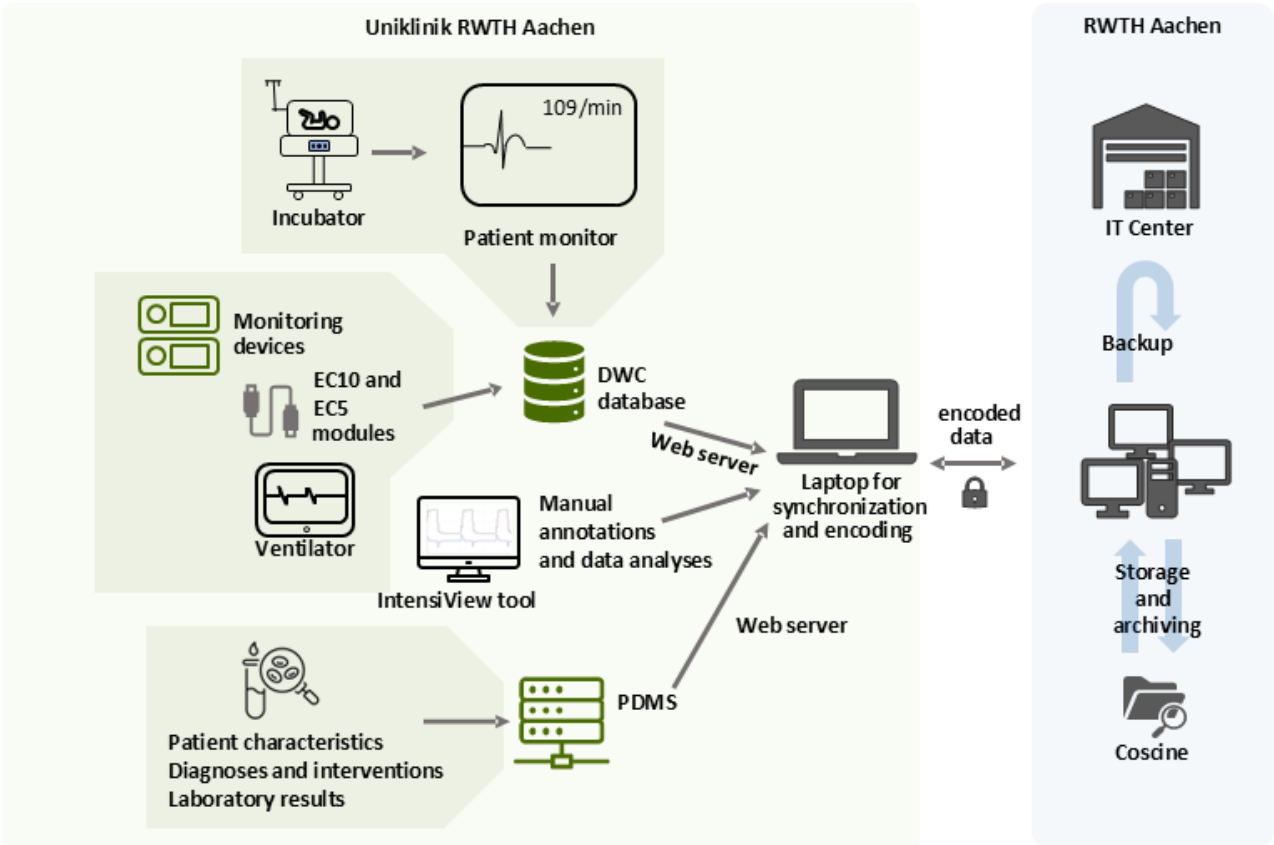
The bedside monitoring is conducted using a Philips IntelliVue X2 monitor (all Philips devices provided by Philips GmbH Market DACH) connected to an IntelliVue MP70 monitor. A Sentec Digital Monitor with V-Sign sensor (SENTEC AG) is used for transcutaneous CO₂ monitoring. Ventilatory support is supplied by a Leoni plus or Elisa ventilator (Löwenstein Medical). The ventilators and CO₂ monitoring are connected to the bedside monitoring using Philips IntelliBridge EC5 ID modules connected via a local area network cable to IntelliBridge EC10 modules. EC10 modules are connected to the monitor by a module rack M8048A. All monitoring and ventilation data are retrieved after storage from the Philips Patient Information Center iX using Philips Data Warehouse Connect, version C.03.

Parameters derived from the PDMS (CGM Medico [CompuGroup Medical SE and Co KGaA]) are extracted

manually and retrospectively added to the dataset. These comprise patient characteristics (eg, admission weight and sex), diagnoses, laboratory results, body positioning, and interventions (imaging, intubations and extubations, operations, and nursing care). Results from manual data analyses are annotated using a project-internal visualization and annotation tool (refer to the IntensiView Tool section). These annotations included patient-ventilator asynchronies, apnea of prematurity, exact time of blood sampling, disconnection of ventilators or monitoring devices, or suctioning.

All data, including timelines, are stored in a coded form in a database hosted by the Chair of Embedded Software. The hospital intern patient identifiers are encoded using a randomly generated 32-character universally unique identifier. The encoding table is stored in an access-protected file at Uniklinik RWTH Aachen and only accessible to clinical investigators. A data flowchart is provided in Figure 1.

Figure 1. Data flowchart. DWC: Data Warehouse Connect; PDMS: patient data management system; RWTH: Rheinisch-Westfälische Technische Hochschule.



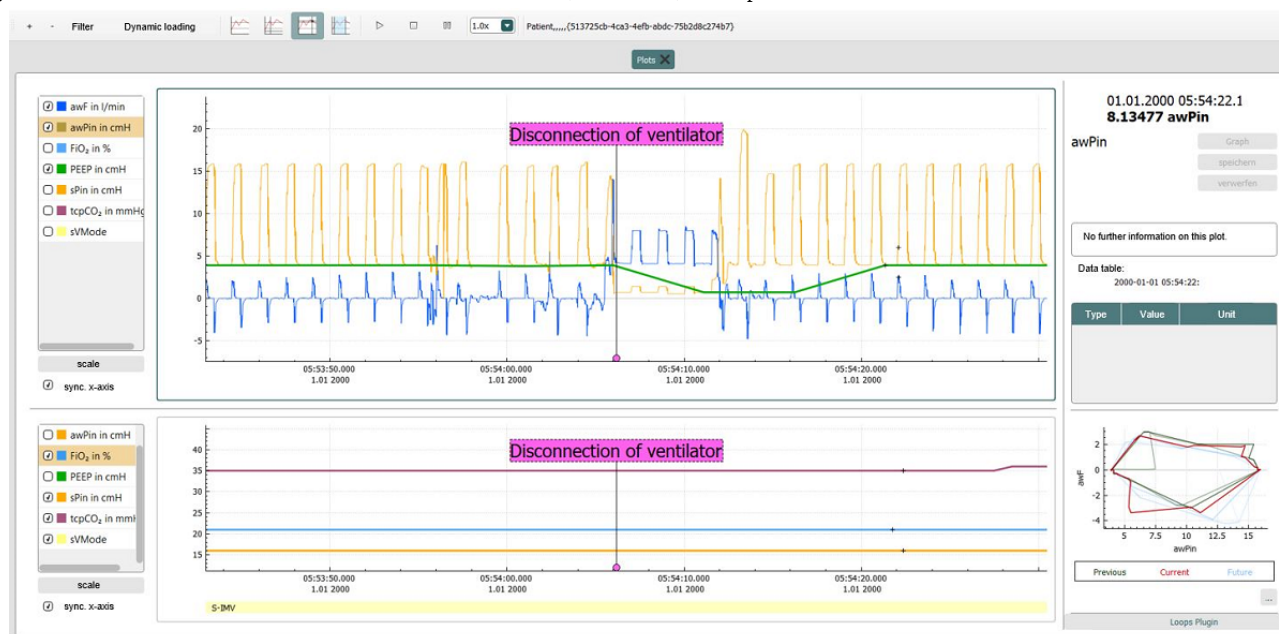
IntensiView Tool

The IntensiView tool (RWTH Aachen University) is a project-specific software developed by the Chair of Embedded Software. The tool is developed in C++ using the Qt Framework to easily deploy across operating systems. A PHP-based backend is used to communicate with the database. IntensiView is used for data synchronization, visualization, annotation, and export. Visualization can be adapted to the user's requirements by changing colors, adapting the number of plots with freely selectable axes, and switching between column and line diagrams. The annotations can then be added manually or automatized using detection algorithms and are inserted into the database. A screenshot is shown in Figure 2.

An annotation search tool enables the user to reload episodes with specific annotations directly. All data can be exported in CSV format. Additionally, external time-series data and annotations can be easily imported and visualized.

A plug-in interface offers standardized access to the data loaded in the IntensiView tool to integrate new functionality, such as detection algorithms that support current research questions. The tool offers a variety of plug-ins for data analyses, for example, an ECG analysis tool, parameter calculation (eg, mean and median, calculation of airway volume curve from the flow curve, and calculation of airway resistance and compliance from ventilator pressure and flow waveform data), and the creation of ventilator loops.

Figure 2. Screenshot of the IntensiView tool for the visualization, annotation, and export of clinical data.



Bias

We included all eligible patients to avoid selection bias. All patients were treated according to local standards. The study participation did not alter diagnostics and treatment. Monitoring and measurements were conducted by the same type of measuring devices in all patients to minimize detection bias.

Study Size

The aim was to include all neonatal and pediatric intensive care patients at Uniklinik RWTH Aachen. Due to limited staff resources, short-stay patients may sometimes not be included because they are discharged before a clinical investigator can link their patient identifiers with the corresponding patient data.

Quantitative Variables

The following variables were stored in a categorized fashion: sex, age or gestational age, admission weight, admission weight percentile, and type of annotation.

Statistical Methods

Descriptive statistics, including medians, IQRs, and frequency distributions, were computed to characterize the dataset. No inferential statistical analyses were conducted. Statistical

analyses were performed using SPSS (version 29.0.0.0; IBM Corp).

Data Access and Cleaning Methods

Only the clinical investigators who were directly involved in patient treatment had access to the linking file with the identifiers and were able to access the decoded data at all times. The other project partners had unlimited access to the coded dataset.

After synchronization, the data were reviewed for completeness and consistency. Data cleaning involved identifying missing or incomplete values, ensuring consistency across variables, and resolving any discrepancies. Data anomalies, such as outliers (eg, z scores >3), were detected manually or by statistical means. Where appropriate, missing data, duplicate entries, and outliers were removed prior to analyses.

Linkage

Data were not linked across multiple databases.

Results

There were 517 ICU admissions during the study period (17 months as of January 27, 2026). Of these, 117 admissions could not be included due to technical issues (primarily Data Warehouse Connect downtime or faulty linkage between intern patient identifiers and Data Warehouse Connect) or missing linkage between patient identifiers and patient data, resulting in a dataset comprising 400 admissions from 359 individual patients.

In total, 227 admissions were attributable to neonatal patients. Of these, 28 were extremely preterm (<28 weeks), 34 very preterm (28+0 to 31+6 weeks), 68 moderate to late preterm (32+0 to 36+6 weeks), and 97 term (≥37 weeks). In total, 173 pediatric admissions were included (34 infants aged <1 year, 45 toddlers aged 1-5 years, 43 children aged 6-12 years, and 51 adolescents aged 13-18 years).

A total of 57.8% (231/400) of admissions were contributed by male study participants. Of the 359 patients, 31 (8.6%) had multiple admissions and 9 (2.5%) had more than 2 admissions. Mortality was 3.6% (14/400) in the study cohort. Median length of stay was 2 (IQR 1-8) days. In total, approximately 3600 patient days were incorporated in the dataset. Patient characteristics are shown in Table 2.

More than 20,000 annotations were added manually to the dataset. The most frequent annotations are active expiration against the ventilator (n=4083, 20.4%), nursing (n=2926, 14.6%), ineffective efforts during expiration (n=1221, 6.1%), apnea of prematurity (n=1271, 6.4%), and time of medical imaging (sonographies: n=185, 0.9% and X-rays: n=153, 0.8%).

So far, the dataset comprises 3 TB of data (implemented using MariaDB as the database management system) and has been used for multiple research questions. A high time resolution was required to address these subjects. An overview is given in Table 3.

Table 2. Patient characteristics.

	Neonatal	Pediatric	Total
Number of admissions, n/N (%)	227/400 (56.8)	173/400 (43.3)	400/400 (100)
Number of patients, n/N (%)	225/359 (62.7)	134/359 (37.3)	359/359 (100)
Number of readmissions, n/N (%)	2/227 (0.9)	39/173 (22.5)	41/400 (10.3)
Gestational age (wk), median (IQR)	35 (31-38)	N/A ^a	N/A
Age (y), median (IQR)	N/A	10.6 (1.2-13.0)	N/A
Male sex, n/N (%)	131/227 (57.7)	100/173 (57.8)	231/400 (57.8)
Length of stay (d), median (IQR)	11.75 (1-12.75)	2 (1-7)	2 (1-8)
Mortality, n/N (%)	10/227 (4.4)	3/173 (1.7)	14/400 (3.6)
Birth weight (g), median (IQR)	1805 (1446-3251)	N/A	N/A
Admission weight (kg), median (IQR)	N/A	22 (11-50)	N/A

^aN/A: not available.

Table 3. Published studies conducted with AIX-Neo-Guard data.

Subject	Data type	Patient group	Number of patients	Number of annotations	Publication
IEE ^a	Airway flow curve	Mechanically ventilated neonates	10	2665	[30]
IEE	Airway flow curve	Mechanically ventilated neonates	10	2665	[31]
Breath classification	Airway flow and pressure curve	Mechanically ventilated neonates	18	6304	[32]
Hypoglycemia and hyperglycemia	- ECG^b - Blood glucose	Neonates	12	251	[33]
Lung model	- Peak inspiratory pressure - PEEP^c - FiO₂^d - Respiratory rate - Blood gas analyses	Neonates	11	460	[34]
tcPCO ₂ ^e	- tcPCO₂ - Blood gas analyses	Neonates	12	172	[35]

^aIEE: ineffective efforts during expiration.

^bECG: electrocardiography.

^cPEEP: positive end-expiratory pressure.

^dFiO₂: inspired oxygen fraction.

^etcPCO₂: transcutaneous CO₂ monitoring.

Discussion

Key Results

In this collaborative project, we successfully established a monocentric high-frequency database of neonatal and pediatric intensive care patients. The dataset incorporates 400 admissions with a total of 3600 treatment days and more than 20,000 manual annotations. We store data from bedside monitoring, ventilation, laboratory results, patient characteristics, and a variety of manual annotations. There is currently no published description of comparable European high-frequency datasets for neonatal and pediatric intensive care patients.

Interpretation

There are various existing databases incorporating neonatal or pediatric ICU data. Many of these databases contain patient-level data but do not provide high-frequency data [36]. The Pediatrix BabySteps Data Warehouse is a large database of more than 1 million neonatal patients [37]. It contains patient data and characteristics recorded on a daily basis. Many of these data are also included in our database. We collect a variety of additional variables, many of which are captured with high temporal resolution at the second or subsecond sampling levels.

Previously published intensive care databases have paved the way for secondary data analyses. The most commonly used datasets are the Medical Information Mart for Intensive Care-III (MIMIC-III) and MIMIC-IV databases [38,39]. These databases contain medical and demographic data from more than 350,000 ICU patients and have been used in more than 2000 secondary data analyses [40]. However, pediatric patients

only constitute a small proportion of the MIMIC-III cohort and were completely excluded from MIMIC-IV. The MIMIC-III Waveform Database [41] extends the clinical data from MIMIC-III by waveform data from approximately 30,000 ICU patients, including 7870 neonates. However, only a subset of these recordings has been matched with the clinical data from MIMIC-III [42]. Comparing the IDs from the matched subset with the original one, we found only 6 neonatal patients with both clinical and waveform data.

Specific pediatric datasets exist: the Pediatric Intensive Care database, for example, incorporates data from more than 12,000 neonatal and pediatric patients [43]. Among others, it provides vital signs, medications, laboratory measurements, and lengths of stay. However, these are stored at low recording frequencies (ie, at 1-minute intervals). The investigation of pathophysiologic conditions often benefits from much higher data resolutions [44].

An example of such a high-frequency dataset is the Preterm Infant Cardio-Respiratory Signals Database [45]. It incorporates 1.6 GB of high-resolution data of ECG and chest impedance from 10 spontaneously breathing premature infants. However, the small sample size and restricted parameter set limit its potential applications.

For pediatric patients, a Canadian group established a database of approximately 1200 patients treated in a pediatric ICU [46]. The researchers collected data at higher frequencies compared to other available databases. The data were recorded every 5 seconds from the bedside monitors and every 30 seconds from the ventilators. This represents a major advancement for research

involving high-frequency data. However, this time resolution does not allow a reconstruction of waveform data such as ECG or the airway pressure waveform. In previous analyses, we developed algorithms for detection of ineffective efforts during expiration. They constitute one of the most prevalent forms of patient-ventilator asynchronies [47] and result in prolonged mechanical ventilation and length of stay [48]. The detection algorithms are based on synchronous breath-by-breath waveform analyses of airway flow, which requires a very high time resolution. To the best of our knowledge, there are no descriptions of neonatal or pediatric datasets that can provide high-frequency data with an equivalent parameter set for a comparable sample size in Europe.

An approach for collecting high-frequency data was developed for The Hospital for Sick Children in Toronto, Canada, establishing a new data collection system for waveform storage and retrieval [49]. The database used (AtriumDB) holds waveforms from more than 5300 patients. The paper focuses on the technical development and compression techniques, giving only a short overview of the available data. Furthermore, only waveform data were stored through the described setup, without additional data relevant for intensive care.

Besides the high-resolution monitoring and ventilation data, the AIX-Neo-Guard dataset incorporates important diagnoses, laboratory results, patient characteristics, and manual annotations. This enables a detailed reconstruction of patient states and dynamic physiological changes. In particular, machine learning approaches benefit from such a dataset. The large number of manual annotations can be used to train detection algorithms for highly relevant complications. Furthermore, the unlabeled data can be used for unsupervised or self-supervised learning approaches to detect anomalies, identify clusters, or train foundation models. The integration of high-frequency physiological signals with clinical variables offers an important platform for clinically meaningful and translational research. The dataset opens important opportunities toward outcome-oriented research and enhances pathophysiologic understanding [50]. For example, using detection methods for ineffective efforts based on waveforms together with information on the patient's state can deepen our understanding of the effects of patient-ventilator asynchrony on the patient's outcome.

Limitations

The most important limitation is the single-center study design. First, it limits the available study population. Furthermore, diagnostic and treatment practices are based on local standards, which may introduce institutional and performance bias. We are preparing to collaborate with other centers to increase the available data and reduce the influence of local practices. Joint research across institutions is also desirable to allow the validation of algorithms (eg, for clinical decision support) on external data. This can contribute to enhancing the generalizability of results derived from a particular dataset.

So far, the dataset does not incorporate all relevant information on diagnostics and treatment. Fluid management and medication, for instance, are not yet integrated into the parameter set. Furthermore, the duration, for example, of nursing care is not labeled. This, in turn, is a current limitation for the possible use

cases of the dataset. Predictive modeling and causal inference often require this valuable information to provide reliable estimates. The legal framework and ethics approval permit the storage of all variables related to diagnostics and treatment in the ICU. We extend the number of collected parameters continuously and plan to incorporate fluid management and medication data in the future.

Due to the rapid growth of data, there are insufficient staff resources to manually annotate the full dataset. The annotations are continuously extended and improved to enhance the quality of data. The set of annotations may furthermore also be enhanced by automated annotation using machine learning approaches. As previous analyses specifically focused on respiratory events, respiratory annotations are overrepresented compared to nonrespiratory annotations. These annotations will have to be made prior to use for nonrespiratory use cases. The data quality is furthermore hampered by technical issues. Occasionally, the ventilator fails to connect to the central monitoring interface. Resolving this issue would require restarting the ventilator. However, during mechanical ventilation, such an intervention would not be ethically justifiable because it could jeopardize the safety of critically ill patients. Similarly, the patient identifiers must be linked to the patient data on the bedside monitor to include the patient in the dataset. The limited staff resources do not allow a clinical investigator to be present at all times to carry out this step. This problem therefore systematically affects short-stay patients who require only short-term treatment or observation (such as those with transient tachypnea of the newborn, those recovering from operative procedures, or those admitted for intoxications). These episodes result in missing valuable data. This introduces a relevant selection bias toward more severe and longer-stay patients and has an impact on the representativeness of the dataset. Specifically, the overrepresentation of infants with longer or more severe courses will influence descriptive analyses and downstream modeling based on this cohort. As a result, findings derived from this dataset should be interpreted as reflecting the subset of neonatal ICU patients for whom complete data capture was feasible, rather than the entire admitted population.

Finally, the dataset was established to promote secondary data analyses. Secondary data research is always prone to bias due to unmeasured confounding and missing data. As the data were collected for a different purpose, the available data do not always meet the requirements of a specific research question, for example, if parameters of interest are missing. In these cases, other data sources must be sought. We hope to reduce this source of bias by pooling data with other centers in the future, increasing the set of available parameters, and continuing data collection and annotation. While secondary data research can support evidence-based medicine, interventional trials are sometimes still necessary.

Future Work

We are currently working on machine learning algorithms to automatically detect and correct time stamps for nursing rounds and blood sampling. These are the first steps toward an

automated improvement of data quality, as these interventions inevitably lead to data anomalies, such as noisy ECG.

To further extend the dataset, information from pregnancy medical history is currently being integrated into the database for all neonates. In the future, the dataset will be further enriched with microbiological data (eg, results from blood cultures) and more comprehensive information on fluid management and medication administration. This will help improve the interpretability and increase the number of research questions that can be addressed using the dataset.

By sharing these database descriptions, we hope to encourage cross-institutional cooperation and invite all interested researchers to contact us, for example, to conduct external validation of algorithms. We hope that collaborations will promote the multicenter pooling of data and thereby enhance the available data. Larger study cohorts and datasets can improve statistical power and are particularly important for investigating rare diseases, for which adequate sample sizes are often difficult to achieve in single-center studies.

Acknowledgments

The authors declare that no generative AI was used.

Funding

The AIX-Neo-Guard project was funded by the German Federal Ministry of Research, Technology, and Space (funding: 031L0303).

Data Availability

Data are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: LSO, MS

Data curation: LSO

Formal analysis: LSO

Funding acquisition: AS, MS

Methodology: LSO

Project administration: AS, MS

Resources: LSO, MS, TO

Software: AS, CLO

Supervision: MS

Validation: CLO

Visualization: LSO

Writing – original draft: LSO

Writing – review and editing: AS, CLO, MS, TO

Conflicts of Interest

None declared.

References

1. Killien EY, Keller MR, Watson RS, Hartman ME. Epidemiology of intensive care admissions for children in the US from 2001 to 2019. *JAMA Pediatr*. May 01, 2023;177(5):506-515. [FREE Full text] [doi: [10.1001/jamapediatrics.2023.0184](https://doi.org/10.1001/jamapediatrics.2023.0184)] [Medline: [36972043](https://pubmed.ncbi.nlm.nih.gov/36972043/)]
2. Braun D, Braun E, Chiu V, Burgos AE, Gupta M, Volodarskiy M, et al. Trends in neonatal intensive care unit utilization in a large integrated health care system. *JAMA Netw Open*. Jun 01, 2020;3(6):e205239. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.5239](https://doi.org/10.1001/jamanetworkopen.2020.5239)] [Medline: [32556257](https://pubmed.ncbi.nlm.nih.gov/32556257/)]

3. Peters MJ, Argent A, Festa M, Leteurtre S, Piva J, Thompson A, et al. The intensive care medicine clinical research agenda in paediatrics. *Intensive Care Med.* Sep 2017;43(9):1210-1224. [doi: [10.1007/s00134-017-4729-9](https://doi.org/10.1007/s00134-017-4729-9)] [Medline: [28315043](#)]
4. Bourgeois FT, Murthy S, Pinto C, Olson KL, Ioannidis JP, Mandl KD. Pediatric versus adult drug trials for conditions with high pediatric disease burden. *Pediatrics.* Aug 2012;130(2):285-292. [FREE Full text] [doi: [10.1542/peds.2012-0139](https://doi.org/10.1542/peds.2012-0139)] [Medline: [22826574](#)]
5. Lehmann B. Reflections on the regulatory field covering the development of paediatric medicinal products: a brief overview of current status and challenges. *Front Pharmacol.* Jun 3, 2024;15:1375988. [FREE Full text] [doi: [10.3389/fphar.2024.1375988](https://doi.org/10.3389/fphar.2024.1375988)] [Medline: [38887557](#)]
6. Landwehr C, Richardson J, Bint L, Parsons R, Sunderland B, Czarniak P. Cross-sectional survey of off-label and unlicensed prescribing for inpatients at a paediatric teaching hospital in Western Australia. *PLoS One.* Jan 8, 2019;14(1):e0210237. [FREE Full text] [doi: [10.1371/journal.pone.0210237](https://doi.org/10.1371/journal.pone.0210237)] [Medline: [30620749](#)]
7. McCrory C, Dooley C, Layte R, Kenny RA. The lasting legacy of childhood adversity for disease risk in later life. *Health Psychol.* Jul 2015;34(7):687-696. [doi: [10.1037/hea0000147](https://doi.org/10.1037/hea0000147)] [Medline: [25150540](#)]
8. Barker DJ, Eriksson JG, Forsén T, Osmond C. Fetal origins of adult disease: strength of effects and biological basis. *Int J Epidemiol.* Dec 2002;31(6):1235-1239. [doi: [10.1093/ije/31.6.1235](https://doi.org/10.1093/ije/31.6.1235)] [Medline: [12540728](#)]
9. Herrera-Carrillo FE, Patel R, Flores S, Villarreal EG, Farias JS, Loomba RS. Randomized controlled trials in pediatric cardiology: a power struggle? *Pediatr Cardiol.* Feb 2023;44(2):306-311. [doi: [10.1007/s00246-022-03039-z](https://doi.org/10.1007/s00246-022-03039-z)] [Medline: [36324012](#)]
10. Cho SM, Serghiou S, Ioannidis JP, Klassen TP, Contopoulos-Ioannidis DG. Large pediatric randomized clinical trials in ClinicalTrials.gov. *Pediatrics.* Sep 2021;148(3):e2020049771. [doi: [10.1542/peds.2020-049771](https://doi.org/10.1542/peds.2020-049771)] [Medline: [34465592](#)]
11. Michelson KA, Cushing AM, Buchholz EM. National trends in pediatric inpatient capacity. *JAMA Pediatr.* Feb 01, 2025;179(2):208-209. [doi: [10.1001/jamapediatrics.2024.5110](https://doi.org/10.1001/jamapediatrics.2024.5110)] [Medline: [39621328](#)]
12. Speer EM, Lee LK, Bourgeois FT, Gitterman D, Hay WWJ, Davis JM, et al. The state and future of pediatric research-an introductory overview: the state and future of pediatric research series. *Pediatr Res.* Jan 24, 2023;1-5. [FREE Full text] [doi: [10.1038/s41390-022-02439-4](https://doi.org/10.1038/s41390-022-02439-4)] [Medline: [36694026](#)]
13. Park SS, Grayson MH. Clinical research: protection of the "vulnerable"? *J Allergy Clin Immunol.* May 2008;121(5):1103-1107. [doi: [10.1016/j.jaci.2008.01.014](https://doi.org/10.1016/j.jaci.2008.01.014)] [Medline: [18313131](#)]
14. Shaikh H, Lyle AN, Oslin E, Gray MM, Weiss EM. Eligible infants included in neonatal clinical trials and reasons for noninclusion: a systematic review. *JAMA Netw Open.* Oct 01, 2024;7(10):e2441372. [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.41372](https://doi.org/10.1001/jamanetworkopen.2024.41372)] [Medline: [39453652](#)]
15. Mayer SL, Brajcich MR, Juste L, Hsu JY, Yehya N. Racial and ethnic disparity in approach for pediatric intensive care unit research participation. *JAMA Netw Open.* May 01, 2024;7(5):e2411375. [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.11375](https://doi.org/10.1001/jamanetworkopen.2024.11375)] [Medline: [38748423](#)]
16. Gitterman DP, Langford WS, Hay WWJ. The uncertain fate of the National Institutes of Health (NIH) pediatric research portfolio. *Pediatr Res.* Sep 2018;84(3):328-332. [doi: [10.1038/s41390-018-0035-7](https://doi.org/10.1038/s41390-018-0035-7)] [Medline: [29976967](#)]
17. Johnston MP. Secondary data analysis: a method of which the time has come. *Qual Quant Methods Libr.* 2014;3(3). [FREE Full text]
18. Rew L, Koniak-Griffin D, Lewis MA, Miles M, O'sullivan A. Secondary data analysis: new perspective for adolescent research. *Nurs Outlook.* 2000;48(5):223-229. [doi: [10.1067/mno.2000.104901](https://doi.org/10.1067/mno.2000.104901)] [Medline: [11044297](#)]
19. Pant D, Nytrø Ø, Leventhal BL, Clausen C, Koochakpour K, Stien L, et al. Secondary use of health records for prediction, detection, and treatment planning in the clinical decision support system: a systematic review. *BMC Med Inform Decis Mak.* May 16, 2025;25(1):190. [FREE Full text] [doi: [10.1186/s12911-025-03021-8](https://doi.org/10.1186/s12911-025-03021-8)] [Medline: [40380138](#)]
20. Lee JH, Byon HJ, Choi S, Jang YE, Kim EH, Kim JT, et al. Safety and efficacy of off-label and unlicensed medicines in children. *J Korean Med Sci.* Jul 19, 2018;33(37):e227. [FREE Full text] [doi: [10.3346/jkms.2018.33.e227](https://doi.org/10.3346/jkms.2018.33.e227)] [Medline: [30190655](#)]
21. Zhong Y, Zhang X, Zhou L, Li L, Zhang T. Updated analysis of pediatric clinical studies registered in ClinicalTrials.gov, 2008-2019. *BMC Pediatr.* Apr 30, 2021;21(1):212. [FREE Full text] [doi: [10.1186/s12887-021-02658-4](https://doi.org/10.1186/s12887-021-02658-4)] [Medline: [33931029](#)]
22. Pollack MM, Holubkov R, Reeder R, Dean JM, Meert KL, Berg RA, et al. PICU length of stay: factors associated with bed utilization and development of a benchmarking model. *Pediatr Crit Care Med.* Mar 2018;19(3):196-203. [FREE Full text] [doi: [10.1097/PCC.0000000000001425](https://doi.org/10.1097/PCC.0000000000001425)] [Medline: [29239978](#)]
23. Varrio T, De Luca D, Kuitunen I. Evidence certainty in neonatology-a meta-epidemiological analysis of Cochrane reviews. *Eur J Pediatr.* Feb 11, 2025;184(2):191. [doi: [10.1007/s00431-025-06023-w](https://doi.org/10.1007/s00431-025-06023-w)] [Medline: [39934466](#)]
24. Zuo Z, Watson M, Budgen D, Hall R, Kennelly C, Al Moubayed N. Data anonymization for pervasive health care: systematic literature mapping study. *JMIR Med Inform.* Oct 15, 2021;9(10):e29871. [FREE Full text] [doi: [10.2196/29871](https://doi.org/10.2196/29871)] [Medline: [34652278](#)]
25. Ganslandt T, Helbing K, Drepper T, Pommerening K. Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0. Berlin, Germany. Medizinisch Wissenschaftliche Verlagsgesellschaft; 2014.

26. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep*. Jul 28, 2020;10(1):12598. [FREE Full text] [doi: [10.1038/s41598-020-69250-1](https://doi.org/10.1038/s41598-020-69250-1)] [Medline: [32724046](https://pubmed.ncbi.nlm.nih.gov/32724046/)]
27. De Luca D, Modi N, Davis P, Kusuda S, de Wildt SN, Keszler M, et al. The Lancet Child and Adolescent Health Commission on the future of neonatology. *Lancet Child Adolesc Health*. Aug 2025;9(8):578-612. [doi: [10.1016/S2352-4642\(25\)00106-3](https://doi.org/10.1016/S2352-4642(25)00106-3)] [Medline: [40580970](https://pubmed.ncbi.nlm.nih.gov/40580970/)]
28. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. Nov 27, 2013;310(20):2191-2194. [doi: [10.1001/jama.2013.281053](https://doi.org/10.1001/jama.2013.281053)] [Medline: [24141714](https://pubmed.ncbi.nlm.nih.gov/24141714/)]
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. Apr 2008;61(4):344-349. [doi: [10.1016/j.jclinepi.2007.11.008](https://doi.org/10.1016/j.jclinepi.2007.11.008)] [Medline: [18313558](https://pubmed.ncbi.nlm.nih.gov/18313558/)]
30. Oprea C, Stegemann L, Olivier LS, Buglowski M, Becker S, Orlikowsky T, et al. Detecting ineffective efforts during expiration for neonates with attention RNNs. *Curr Dir Biomed Eng*. Dec 19, 2024;10(4):478-481. [doi: [10.1515/cdbme-2024-2117](https://doi.org/10.1515/cdbme-2024-2117)]
31. Oprea C, Kannan A, Olivier L, Schoberer M, Stollenwerk A. Comparing a rule-based decision tree to an LSTM network for the detection of ineffective efforts during expiration in neonates. *Curr Dir Biomed Eng*. Sep 12, 2025;11(1):370-373. [doi: [10.1515/cdbme-2025-0194](https://doi.org/10.1515/cdbme-2025-0194)]
32. Oprea C, Grüne M, Buglowski M, Olivier L, Orlikowsky T, Kowalewski S, et al. Evaluating the explainable AI method Grad-CAM for breath classification on newborn time series data. *IFAC PapersOnLine*. 2024;58(24):123-128. [doi: [10.1016/j.ifacol.2024.11.023](https://doi.org/10.1016/j.ifacol.2024.11.023)]
33. Olivier L, Strunk J, Oprea C, Stollenwerk A, Orlikowsky T, Schoberer M. Analysis of ECG characteristics during phases of hypoglycemia and hyperglycemia in the neonatal period [Article in German]. *Z Geburtshilfe Neonatol*. 2025;229(06):S289-S290. [doi: [10.1055/s-0045-1814062](https://doi.org/10.1055/s-0045-1814062)]
34. Oprea C, Walus N, Buglowski M, Olivier L, Schoberer M, Stollenwerk A. Evaluating the suitability of a neonatal respiratory system model for the simulation of arterial partial pressures. *Curr Dir Biomed Eng*. Sep 12, 2025;11(1):417-420. [doi: [10.1515/cdbme-2025-0206](https://doi.org/10.1515/cdbme-2025-0206)]
35. Olivier L, Oprea C, Pfannschmidt V, Stollenwerk A, Orlikowsky T, Schoberer M. Correlation between tcpCO₂ and capillary pCO₂ as a function of gestational age [Article in German]. *Z Geburtshilfe Neonatol*. 2025;229(03):e12. [doi: [10.1055/s-0045-1808326](https://doi.org/10.1055/s-0045-1808326)]
36. Statnikov Y, Ibrahim B, Modi N. A systematic review of administrative and clinical databases of infants admitted to neonatal units. *Arch Dis Child Fetal Neonatal Ed*. May 2017;102(3):F270-F276. [FREE Full text] [doi: [10.1136/archdischild-2016-312010](https://doi.org/10.1136/archdischild-2016-312010)] [Medline: [28087722](https://pubmed.ncbi.nlm.nih.gov/28087722/)]
37. Spitzer AR, Ellsbury D, Clark RH. The Pediatrix BabySteps® Data Warehouse--a unique national resource for improving outcomes for neonates. *Indian J Pediatr*. Jan 2015;82(1):71-79. [doi: [10.1007/s12098-014-1585-2](https://doi.org/10.1007/s12098-014-1585-2)] [Medline: [25319813](https://pubmed.ncbi.nlm.nih.gov/25319813/)]
38. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. May 24, 2016;3:160035. [FREE Full text] [doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35)] [Medline: [27219127](https://pubmed.ncbi.nlm.nih.gov/27219127/)]
39. Johnson AE, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. Jan 03, 2023;10(1):1. [FREE Full text] [doi: [10.1038/s41597-022-01899-x](https://doi.org/10.1038/s41597-022-01899-x)] [Medline: [36596836](https://pubmed.ncbi.nlm.nih.gov/36596836/)]
40. Zhang X, Kong W, Shi R, Sun L, Xu M, Gong L. Data-driven trends in critical care informatics: a bibliometric analysis of global collaborations using the MIMIC database (2004-2024). *Comput Biol Med*. Sep 2025;195:110670. [doi: [10.1016/j.combiomed.2025.110670](https://doi.org/10.1016/j.combiomed.2025.110670)] [Medline: [40580617](https://pubmed.ncbi.nlm.nih.gov/40580617/)]
41. Moody B, Villarreal M, Clifford GD, Silva I. MIMIC-III Waveform Database (version 1.0). *Physionet*. Apr 07, 2020. [doi: [10.13026/c2607m](https://doi.org/10.13026/c2607m)]
42. Moody B, Moody G, Villarreal M, Clifford GD, Silva I. MIMIC-III Waveform Database Matched Subset (version 1.0). *PhysioNet*. Apr 7, 2020. [doi: [10.13026/c2294b](https://doi.org/10.13026/c2294b)]
43. Zeng X, Yu G, Lu Y, Tan L, Wu X, Shi S, et al. PIC, a paediatric-specific intensive care database. *Sci Data*. Jan 13, 2020;7(1):14. [FREE Full text] [doi: [10.1038/s41597-020-0355-4](https://doi.org/10.1038/s41597-020-0355-4)] [Medline: [31932583](https://pubmed.ncbi.nlm.nih.gov/31932583/)]
44. Zhou W, Chan YE, Foo CS, Zhang J, Teo JX, Davila S, et al. High-resolution digital phenotypes from consumer wearables and their applications in machine learning of cardiometabolic risk markers: cohort study. *J Med Internet Res*. Jul 29, 2022;24(7):e34669. [FREE Full text] [doi: [10.2196/34669](https://doi.org/10.2196/34669)] [Medline: [35904853](https://pubmed.ncbi.nlm.nih.gov/35904853/)]
45. Gee AH, Barbieri R, Paydarfar D, Indic P. Predicting bradycardia in preterm infants using point process analysis of heart rate. *IEEE Trans Biomed Eng*. Sep 2017;64(9):2300-2308. [FREE Full text] [doi: [10.1109/TBME.2016.2632746](https://doi.org/10.1109/TBME.2016.2632746)] [Medline: [27898379](https://pubmed.ncbi.nlm.nih.gov/27898379/)]
46. Brossier D, El Taani R, Sauthier M, Roumeliotis N, Emeriaud G, Juvet P. Creating a high-frequency electronic database in the PICU: the perpetual patient. *Pediatr Crit Care Med*. Apr 2018;19(4):e189-e198. [doi: [10.1097/PCC.0000000000001460](https://doi.org/10.1097/PCC.0000000000001460)] [Medline: [29406373](https://pubmed.ncbi.nlm.nih.gov/29406373/)]
47. Holanda MA, Vasconcelos RD, Ferreira JC, Pinheiro BV. Patient-ventilator asynchrony [Article in English, Portuguese]. *J Bras Pneumol*. 2018;44(4):321-333. [FREE Full text] [doi: [10.1590/S1806-37562017000000185](https://doi.org/10.1590/S1806-37562017000000185)] [Medline: [30020347](https://pubmed.ncbi.nlm.nih.gov/30020347/)]

48. Kyo M, Shimatani T, Hosokawa K, Taito S, Kataoka Y, Ohshimo S, et al. Patient-ventilator asynchrony, impact on clinical outcomes and effectiveness of interventions: a systematic review and meta-analysis. *J Intensive Care*. Aug 16, 2021;9(1):50. [FREE Full text] [doi: [10.1186/s40560-021-00565-5](https://doi.org/10.1186/s40560-021-00565-5)] [Medline: [34399855](https://pubmed.ncbi.nlm.nih.gov/34399855/)]
49. Goodwin AJ, Eytan D, Greer RW, Mazwi M, Thommandram A, Goodfellow SD, et al. A practical approach to storage and retrieval of high-frequency physiological signals. *Physiol Meas*. Apr 20, 2020;41(3):035008. [doi: [10.1088/1361-6579/ab7cb5](https://doi.org/10.1088/1361-6579/ab7cb5)] [Medline: [32131060](https://pubmed.ncbi.nlm.nih.gov/32131060/)]
50. Cesare M, D'Agostino F, Nurchis MC, Magliozzi E, Damiani G, Nursing and Public Health Group, et al. Determinants of prolonged hospitalization in children and adolescents: a retrospective observational study. *J Nurs Scholarsh*. May 2025;57(3):412-430. [doi: [10.1111/jnu.13045](https://doi.org/10.1111/jnu.13045)] [Medline: [39803927](https://pubmed.ncbi.nlm.nih.gov/39803927/)]

Abbreviations

ECG: electrocardiogram

ICU: intensive care unit

MIMIC-III: Medical Information Mart for Intensive Care-III

PDMS: patient data management system

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by M Balcarras; submitted 27.Feb.2026; peer-reviewed by M Görges, M Cesare; comments to author 20.Mar.2026; revised version received 05.Jun.2026; accepted 08.Jun.2026; published 31.Jul.2026

Please cite as:

Olivier LS, Lauterbach Oprea C, Stollenwerk A, Orlikowsky T, Schoberer M

Feasibility of a German High-Frequency Neonatal and Pediatric Intensive Care Dataset (Alx-Neo-Guard Dataset): Secondary Data Cohort Study

JMIR Pediatr Parent 2026;9:e94280

URL: <https://pediatrics.jmir.org/2026/1/e94280>

doi: [10.2196/94280](https://doi.org/10.2196/94280)

PMID: [42536404](https://pubmed.ncbi.nlm.nih.gov/42536404/)

©Lena Sophie Olivier, Camelia Lauterbach Oprea, André Stollenwerk, Thorsten Orlikowsky, Mark Schoberer. Originally published in *JMIR Pediatrics and Parenting* (<https://pediatrics.jmir.org/>), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Pediatrics and Parenting*, is properly cited. The complete bibliographic information, a link to the original publication on <https://pediatrics.jmir.org/>, as well as this copyright and license information must be included.